COUNTS-IN-CELLS
WITH THE
DARK ENERGY SURVEY

BY

ANA ISABEL SALVADOR JUNCO

PHD THESIS IN THEORETICAL PHYSICS

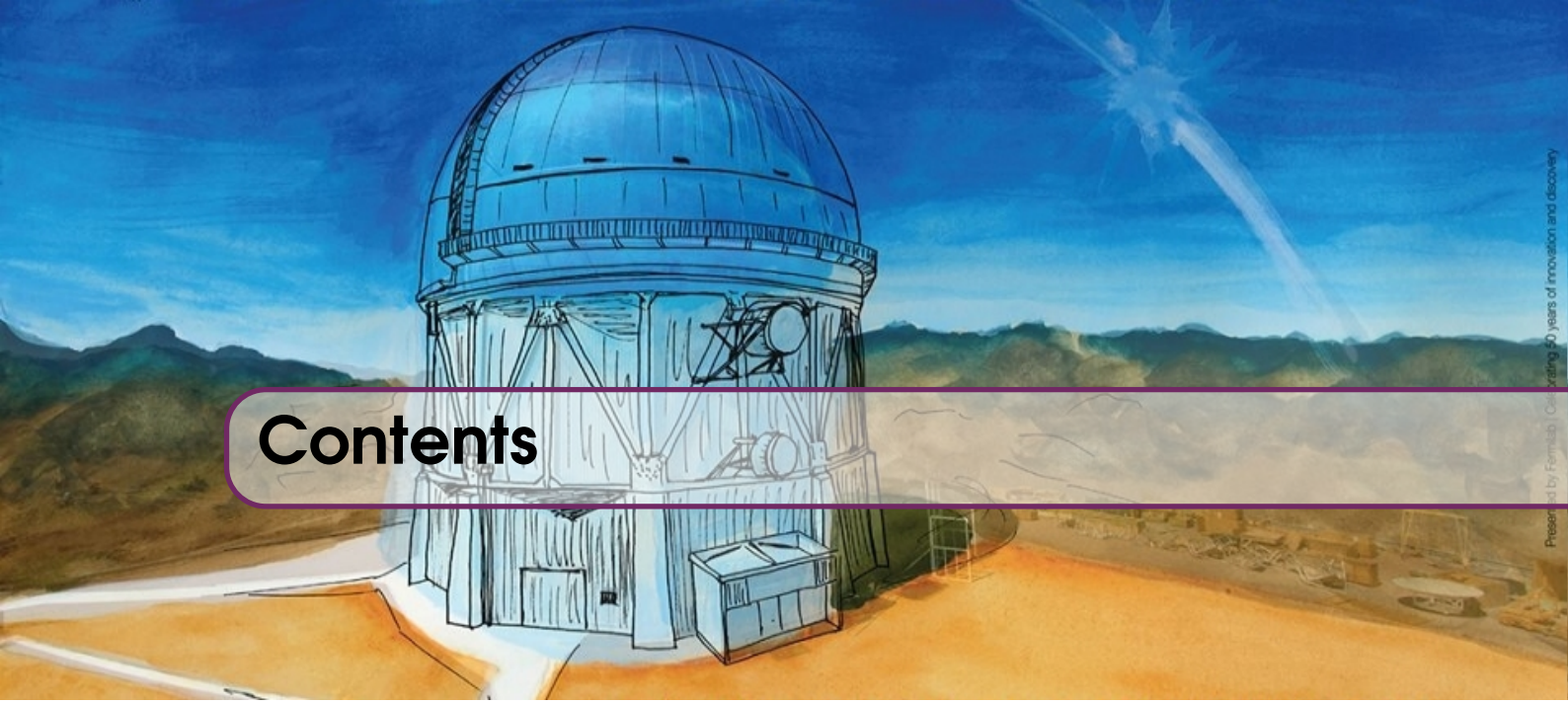
July 12, 2017

SUPERVISED BY

JUAN GARCÍA-BELLIDO CAPDEVILA

UNIVERSIDAD AUTÓNOMA DE MADRID
&
INSTITUTO DE FÍSICA TEÓRICA (UAM-CSIC)

A mi madre



Contents

Authorship

This thesis contains results based on original work done by myself in collaboration with other researchers during my Ph.D. from September 2014 to June 2017. This thesis is done with data and simulations from the 'Dark Energy Survey'.

The summary and the conclusions are written in spanish and english. The chapters 1 and 2 are review chapters about cosmology and large scale structure, based on the literature I have read during my Ph.D.. No new result is presented in these chapters, and no originality is claimed. The chapter 3 is a description of the 'Dark Energy Survey'. And the chapter 4 presents the results of the method developed during my Ph.D. to obtain the linear and non-linear bias from 'Counts-in-Cells'. First we present the results on MICE simulation and then on the Science Verification data from DES. These results are based on an unpublished paper:

- A.I. Salvador, F. J. Sánchez, A. Pagul, J. García-Bellido, E. Sánchez, A. Pujol, J. A. Frieman; 'Counts-in-Cells in the Galaxy Distribution from the Dark Energy Survey Science Verification Data'; (in prep.)

During my Ph.D. I have also been working on BAO, the homogeneity scale and photometric redshifts with genetic algorithms, which led to the published and unpublished papers:

- D. Alonso, A.I. Salvador, F.J. Sánchez, M. Bilicki, J.García-Bellido, E. Sánchez; *'Homogeneity and isotropy in the 2MASS Photometric Redshift catalogue'*; MNRAS.

- A.I. Salvador; M.Manera; 'A Genetic Algorithm for Photometric Redshift Estimation for DES and PAU'; (in prep.)

Agradecimientos

Muchas gracias a todos los que me habéis apoyado durante mi tesis. Sobre todo al final, ha sido duro y han sido muchas las lágrimas. He descubierto que la investigación no es lo mío. Creo que hay que tener un carácter y una paciencia que yo, desde luego, no tengo. Hay que saber afrontar una derrota diaria sin frustrarse, y yo, no he sido capaz. Por otro lado, he descubierto que la docencia me gusta mucho y me da mucha satisfacción.

Primero quiero agradecer a mis padres y mi hermano por haber estado siempre apoyándome y animándome a seguir hacia adelante. Sobre todo, en esos momentos en los que lo único que necesitas es bajarte del mundo. También al resto de mi familia, que aunque nos veamos poco también estáis ahí cuando hace falta.

Muchas gracias a mis amigüitos, Pablo, Irene, Aroa, Dani, Enol, Barry y Valen. Aunque os haya quitado muchos amigüipuntos, son muchos más los que habeis ganado aguantándome. Muchas gracias por los viajes, las noches de fiesta, por estar siempre ahí y por todas las experiencias que hemos compartido.

Muchas gracias también a mis amigas, Carmen, Alba, Blanca, Virginia, Sara, Nerea y Lisa. Gracias por haber estado siempre ahí, por todas las locuras que hemos compartido, por cada vez que nos juntamos y recordamos viejos tiempos.

Muchas gracias a Mab, Víctor y Sara, no sé qué hubiera sido de mí sin nuestras cenas y vuestras tonterías. Muchas gracias a los de la tarde, Andrea, Juan, Ana y Guille. Me han encantado nuestros viajes, tenemos que hacer más!. Muchas gracias a Ana, ha sido genial reencontrate en Madrid después de tantos años.

Muchas gracias a Consuelo, Rocío, Pili y Fran por acogerme desde el primer momento.

Muchas gracias a Juan por haber sido mi director de tesis y a la vez aguantar mis lágrimas. Muchas gracias a Eusebio, por haber respondido siempre a mis dudas y preguntas. Entre los dos habéis sido una combinación entre teoría y práctica que ha hecho que pudiera terminar esta tesis.

Muchas gracias a Javi por enseñárme todo lo que sabes, o intentárlo, y por la paciencia. Muchas gracias a Amanda por lo bien que lo pasámos aquel verano en España. Muchas gracias a David Alonso por ayudarme a empezar.

Muchas gracias a toda la gente del IFT. Muchas gracias a Víctor por todas las conversaciones y pirópos; a Santi por todos los viajes y excursiones (siempre seguiré midiendo las distancias en cintas); a Ander, Xabi y Josu por todas las partidas de cartas; a Sjoerd por ser un buen compi de despacho; y a Aitor porque sin duda es el que más lágrimas ha tenido que aguantar. Muchas gracias a las secretarias por tener siempre una sonrisa y a Isabel por ser mi mami del IFT.

Muchas gracias a toda la gente de DES con la que he compartido grandes momentos en las reuniones de la colaboración. Especialmente Matt, el inglés más español que he conocido nunca. Muchas gracias a Anna, Anais y Manu que habéis hecho los viajes de trabajo más divertidos.

Y por último muchas gracias a Ángel, por tener las mismas ganas de vivir que yo, y por ser capaz de bailar bajo la lluvia. Te quiero.

Resumen

- Timón...
- Sí?
- Te has preguntado alguna vez qué son esos puntitos brillantes del cielo?
- Pumba, no me lo pregunto, lo sé.
- Ah, y qué son?
- Luciérnagas, luciérnagas que se quedaron atrapadas en ese techo azul y negro de arriba.
- Ah, vaya, siempre creí que eran bolas de gas quemándose a millones de kilómetros de aquí.
- Pumba, todo lo tuyo es gas.

(El Rey León)

El descubrimiento de la materia oscura y la energía oscura en el siglo XX ha supuesto una revolución en cosmología. La naturaleza de estos dos componentes del Universo es uno de los grandes retos de la cosmología hoy en día. Desconocemos el origen del 95% del contenido energético del Universo, al 5% que conocemos (la materia que forma los planetas, las estrellas y las galaxias) es lo que llamamos materia bariónica. Las observaciones nos dicen que el $\sim 26\%$ del Universo es materia oscura, un tipo de materia que no vemos pero que sabemos que está ahí por sus efectos gravitatorios. Y por último el $\sim 69\%$ del Universo es energía oscura, un tipo de energía que desconocemos y hace que el Universo se esté expandiendo aceleradamente.

Durante años la cosmología se ha limitado a ser una ciencia principalmente teórica, pero con el desarrollo de la tecnología hemos pasado de ver unas pocas galaxias en placas fotográficas hace 50 años a tener cámaras muy potentes en grandes telecopios que toman imágenes de millones de galaxias hoy en día. Vivimos una época dorada de la cosmología con un creciente número de observaciones de cada vez mayor precisión. Es a través de las observaciones como podemos contrastar las teorías y modelos. Esta época dorada empezó con la comprobación de que el Universo es euclideo y con el descubrimiento de la aceleración del Universo. Desde entonces, cartografiados muy precisos de las anisotropias del fondo cósmico de microondas (CMB), las supernovas de tipo Ia, la distribución espacial

de galaxias y las lentes gravitacionales han confirmado la aceleración del Universo y han llevado a establecer el módelo estándar de la cosmología, llamado ΛCDM por sus siglas 'Cold Dark Matter' o 'Materia Oscura Fría' y Λ por la constante cosmológica.

Hoy en día grandes colaboraciones como Planck, el Sloan Digital Sky Survey (SDSS) o el Dark Energy Survey (DES), toman datos para seguir comprobando el modelo estándar a grandes y pequeñas escalas, e intentar averiguar la naturaleza de la materia oscura y la energía oscura. La colaboración internacional Dark Energy Survey fue creada para descubrir la naturaleza de la energía oscura, y para ello lleva a cabo un cartografiado fotométrico. Esta colaboración ha construído una potente cámara digital de 570 Megapixeles, DECam, instalada en el telescopio Blanco de 4 metros en Cerro Tololo en Chile. Lo que tiene de especial este cartografiado es que combina cuatro sondas de energía oscura: supernovas de tipo Ia, oscilaciones acústicas de bariones, cúmulos de galaxias y lentes gravitacionales. A pesar de ser un cartografiado optimizado para caracterizar la energía oscura, ya se han obtenido resultados sorprendentes más allá del objetivo principal que es la cosmología.

Lo que observamos con un telescopio son galaxias, no materia oscura, ya que ésta no emite luz. Las galaxias se forman en los halos de materia oscura, y por lo tanto trazan la distribución de materia oscura. La relación entre la distribución de galaxias y la de materia es lo que llamamos bias de las galaxias. Entender la relación entre galaxias y materia oscura es esencial para calcular los parámetros del modelo estándar y conocer la distribución real de materia en el Universo.

El método Counts-in-Cells (CiC) o 'Cuenta-por-Celdas' puede utilizarse para calcular el bias de las galaxias. CiC es un método basado en dividir el cielo en celdas de igual volumen y contar el número de galaxias en cada celda. A partir de esto se puede calcular el contraste de densidad de cada celda y por lo tanto los momentos de la distribución del contraste de densidad. A partir de la varianza de la distribución obtenemos el bias lineal, y a partir de los momentos de orden tres y cuatro obtenemos los primeros órdenes del bias no lineal. A lo largo de la tesis se ha desarrollado un método para calcular el bias de las galaxias a partir de CiC. Se ha comprobado que el método funciona con simulaciones y lo he aplicado a datos de la colaboración Dark Energy Survey a la que pertenezco. La principal ventaja de este método, comparado con otros, es que es un método simple y rápido. CiC permite calcular órdenes superiores a la función de correlación a dos puntos sin ser muy costoso computacionalmente.

Summary

- Hey, Timon, ever wonder what those sparkly dots are up there?
- Yes?
- Pumba, I don't wonder; I know.
- Oh. What are they?
- They're fireflies. Fireflies that, uh... got stuck up on that big bluish-black thing.
- Oh, gee. I always thought they were balls of gas burning billions of miles away.
- Pumba, with you, everything's gas.

(The Lion King)

The discovery of dark matter and dark energy in the twentieth century has been a revolution in cosmology. The origin of these two components is completely unknown and it is one of the biggest challenges in cosmology nowadays. We do not know the origin of the 95% of the Universe, the 5% that we do know (what makes up planets, stars and galaxies) is the so-called baryonic matter. Observations suggest that the $\sim 26\%$ of the Universe is dark matter, matter that we cannot see but that we know it is there due to gravitational effects. And the $\sim 69\%$ left is dark energy, an unknown type of energy that makes the Universe to expand in an accelerated way.

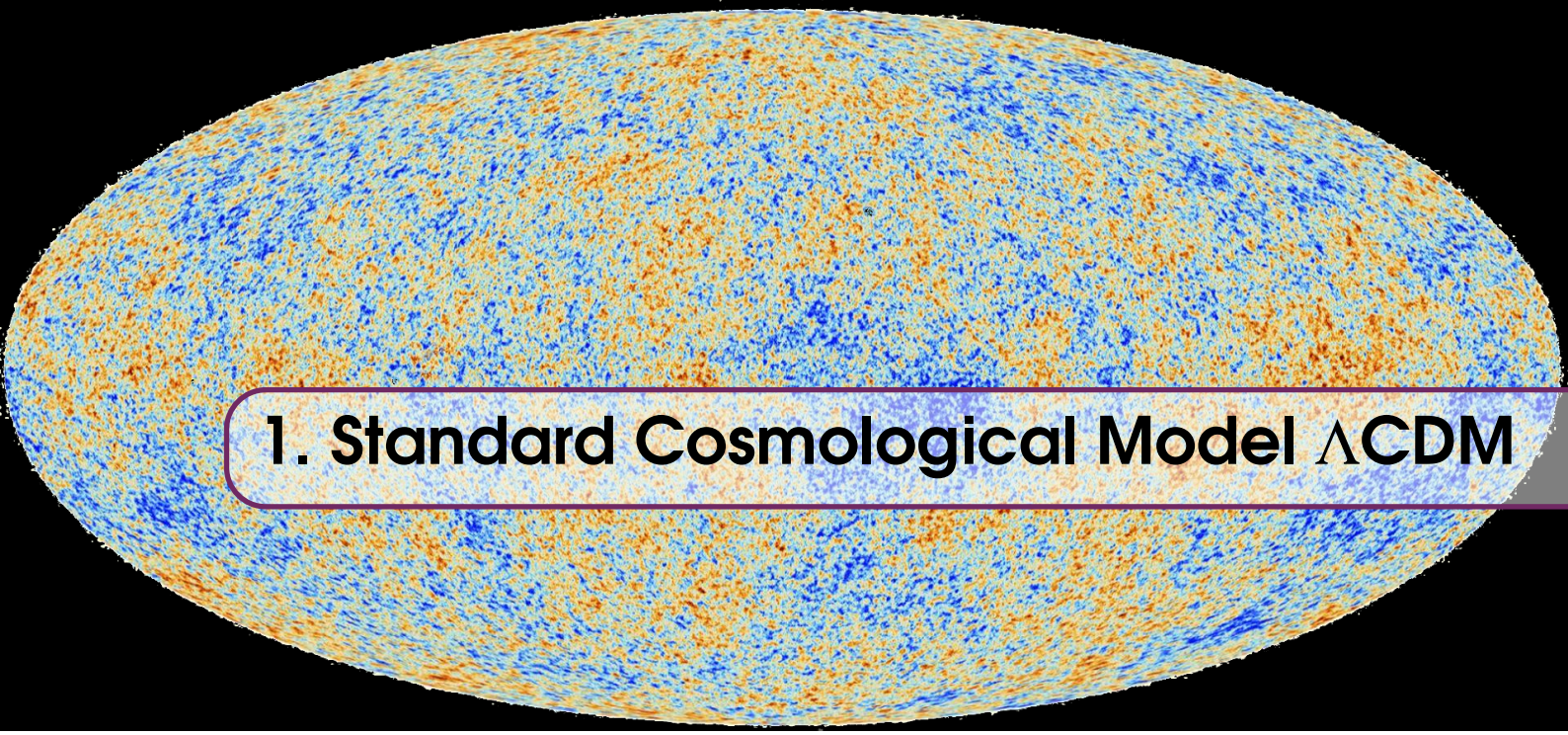
For years cosmology has been a theoretical science, but thanks to technological developments we have moved from observing just a few galaxies in photography plates fifty years ago to the extremely powerful cameras on big telescopes taking images of millions of galaxies nowadays. We are now in a golden age of cosmology, with a growing array of astrophysical observations of ever increasing precision. It is through cosmological observations that we are able to contrast theories and models. This golden age started with the discovery that the Universe is spatially flat and the discovery of the acceleration of the Universe. Since then we have had precise mapping surveys of galaxies and observations of the anisotropies of the cosmic microwave background that have allowed us to know about the Universe. Observations of the Cosmic Microwave Background (CMB), the supernovae Ia, the spatial distribution of galaxies and gravitational lensing have confirmed the acceleration of the Universe and led to the standard cosmological model, called ΛCDM,

acronym of 'Cold Dark Matter' and Λ for the cosmological constant.

Nowadays big international collaborations like Planck, the Sloan Digital Sky Survey (SDSS) or the Dark Energy Survey (DES) are taking data to keep checking the standard model at small and large scales, and try to find out what dark matter and dark energy are. The Dark Energy Survey is a photometric survey created to help uncover the nature of dark energy. This international collaboration built an extremely powerful 570 Megapixel camera, DECam, mounted on the Blanco-4-meter telescope in Chile. What makes DES special is that it combines four probes of dark energy in a single experiment: Type Ia Supernovae, Baryon Acoustic Oscillations, Galaxy clusters and Weak Gravitational Lensing. Although optimized to measure and characterise dark energy, DES has already yielded both expected and unexpected discoveries beyond its primary goal for cosmological studies.

What we observe with the telescopes are galaxies, not dark matter, as it does not interact with light. Galaxies are formed in dark matter halos, hence they can trace the dark matter distribution. The relation between the galaxy and the dark matter distributions is called galaxy bias. Understanding the relation between galaxies and matter is essential for the measurements of cosmological parameters and knowing the real matter distribution in the Universe.

We use the Counts-in-Cells (CiC) method to obtain the galaxy bias. CiC is a method based on dividing the sphere in cells of the same volume and counting the number of galaxies in each cell. Having this, we can compute the density contrast in each cell and then compute the moments of the density contrast distribution. From the variance we can compute the linear bias, and from the third and fourth order moments the first non-linear bias parameters. In this thesis I develop a method to compute the galaxy bias using CiC. First I check that the method works with simulations and then I apply the method to data from the Dark Energy Survey, to which i belong. The best thing about this method, compared to others, It is that it is simple and fast. CiC allows the computation of higher orders than the two-point correlation function without being computationally demanding.



1. Standard Cosmological Model ΛCDM

Modern Cosmology is built according to the Standard Cosmological Model, what we call the ΛCDM model. Thanks to the progress in technology in the last decades, we live in a golden age of cosmology. We can measure the parameters of the Standard Model within small uncertainties, of just a few percent. This model fits a host of astronomical data, however it requires novel physics to explain cosmic acceleration, dark matter, and the origin of the Universe. Further, as our data improves, intriguing discrepancies are starting to arise [1]. Upcoming experiments are the key to clarify the properties of dark matter, dark energy and test our Standard Model [2].

According to Planck's latest results [3], assuming a ΛCDM model, the Universe is formed of a $\sim 70\%$ of dark energy in the form of a cosmological constant and a $\sim 30\%$ of matter, of which the $\sim 84\%$ is dark matter, Figure 1.1. This model is based on the hot Big Bang theory and the Inflationary Paradigm. The Big Bang theory, in turn, is based on the theoretical framework of General Relativity (GR) and four observational facts: the expansion of the Universe, the relative abundance of light elements, the Cosmic Microwave Background (CMB), and the Large-Scale Structure (LSS) in the distribution of galaxies. Although this model describes the evolution of the Universe with precision, it is still incomplete. Nowadays we just know that the dark matter is cold and the dark energy acts as a negative pressure fluid which makes the Universe expand in an accelerated way.

In this chapter I will describe the ΛCDM model. First, I will describe the theoretical framework of the model: General Relativity, Inflation and the Big Bang theory (sec:1.1). Afterwards, I will describe its observational basis (sec:1.2), first the main observations on which the ΛCDM model is based (Supernovae Type Ia, Relative abundance of light elements and the CMB), then other probes of dark energy (Baryon Acoustic Oscillations, Weak Lensing, Redshift Space Distortions and Galaxy Clusters). Furthermore, a current status of the search for dark matter is shown. Finally, a discussion of current and future status of the model and observations is detailed (sec:1.3 and 1.4).

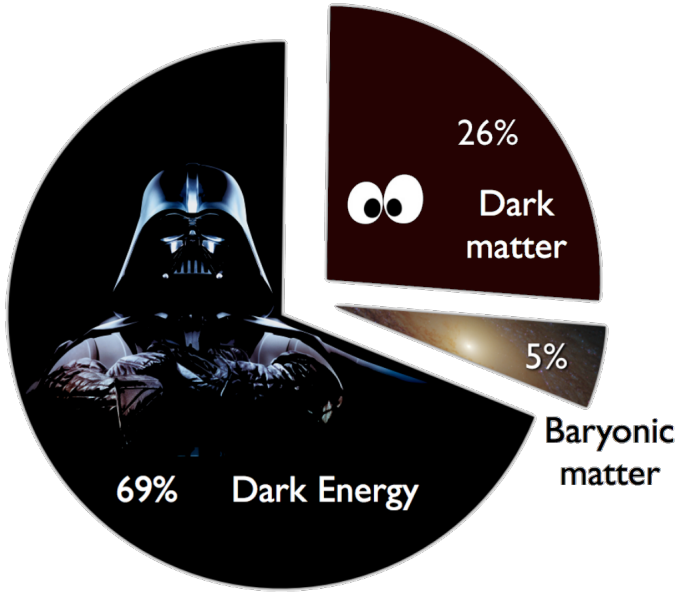


Figure 1.1: Content of the Universe according to Planck's latest results [3], assuming a ΛCDM model. It is formed of a $\sim 69\%$ of dark energy, a $\sim 26\%$ of dark matter, and $\sim 5\%$ of baryonic matter.

1.1 Theoretical framework of the Standard Model

1.1.1 General Relativity (GR)

The development of Modern Cosmology as a quantitative science begun in the 1920s with Einstein's General Relativity [4]. This theory is based on:

- The covariance principle: the invariance of the form of physical laws in different coordinate systems.
- The equivalence principle: Special Relativity laws apply locally for every free falling inertial observer.

GR states that the geometry of space-time is determined by the energy content of the Universe [5]:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{1.1}$$

where $R_{\mu\nu}$ is the Ricci tensor, R is the scalar curvature, Λ is the cosmological constant, $T_{\mu\nu}$ is the energy momentum tensor, G is Newton's gravitational constant, c is the speed of light in vacuum, and $g_{\mu\nu}$ is the metric tensor (with signature $(-,+,+,+)$). These non-linear equations can be solved applying the symmetries of the Universe. Modern Cosmology is based on what is called the Cosmological Principle, all positions and directions in the Universe are equivalent. Although at small scales the Universe looks very inhomogeneous, the deepest galaxy catalogues and the CMB suggest that the Universe is very homogeneous and isotropic on large scales (beyond supercluster scales). The most general metric

satisfying these conditions is the Friedmann-Robertson-Walker (FRW) metric:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-Kr^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] \tag{1.2}$$

written in terms of the invariant geodesic distance $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$, where $a(t)$ is the scale factor, which determines the physical size of the Universe, and K is the spatial curvature of the Universe (K=-1 for an open Universe, K=0 for a flat Universe and K=+1 for a closed Universe).

The most general fluid consistent with the assumption of homogeneity and isotropy is a perfect fluid. This kind of fluid can be completely characterized by its rest frame mass density ρ, and its isotropic pressure p. The energy momentum associated with such a fluid is

$$T_{\mu\nu} = pg_{\mu\nu} + (p+\rho)U_\mu U_\nu \tag{1.3}$$

where U^μ is the comoving four-velocity.

The equations of motion of an observer comoving with this fluid in an expanding Universe can be deduced from Einstein equations (1.1) by substituting the FRW metric (1.2) and the perfect fluid tensor (1.3). These gives us the Friedmann equations:

$$\begin{aligned}\left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{K}{a^2},\\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}(\rho+3p) + \frac{\Lambda}{3}.\end{aligned} \tag{1.4}$$

The Universe will evolve according to its dynamics, and thus the matter/energy content of the Universe.

The rate of expansion of the Universe can be written in terms of the contribution to the energy density of its different components, radiation Ω_R, matter Ω_M, cosmological constant Ω_Λ and curvature Ω_K:

$$H^2(a) = H_0^2\left(\Omega_R\frac{a_0^4}{a^4} + \Omega_M\frac{a_0^3}{a^3} + \Omega_\Lambda + \Omega_K\frac{a_0^2}{a^2}\right) \tag{1.5}$$

where a_0 is the scale factor nowadays and $H_0 = \dot{a}/a(t_0)$ is the Hubble constant. This constant is usually written as $H_0 = 100h$ km s^{-1}Mpc^{-1}, where $h = 0.6774 \pm 0.0046$ [3]. We can rewrite Friedmann equation today ($a = a_0$) as a cosmic sum rule:

$$1 = \Omega_M + \Omega_\Lambda + \Omega_K \tag{1.6}$$

where the contribution of relativistic particles to the total density of the Universe nowadays can be neglected ($\Omega_{CMB} \sim 2.4 \times 10^{-5}h^{-2}$).

The cosmological redshift is a direct consequence of the Hubble expansion, determined by $a(t) = \frac{1}{1+z}$. The frequency of an emitted light ν_1 would be seen redshifted ν_2 by a distant observer:

$$z \equiv \frac{\nu_1 - \nu_2}{\nu_2} \tag{1.7}$$

Initially Einstein introduced a cosmological constant to allow time-independence, spatially homogeneous matter density and constant positive space curvature. Although Einstein did not frame it this way, one can view this constant Λ as representing a constant energy density of the vacuum [6], whose repulsive gravitational effect balances the attractive gravity of matter and thereby allows a static solution. After the discovery of cosmic expansion [7], the cosmological term appeared unnecessary, and Einstein and de Sitter discarded the constant. In the late 1990s, supernova surveys by two independent teams provided direct evidence for accelerating cosmic expansion today [8, 9], and again the cosmological constant was introduced, but now to allow an accelerated expansion. The theoretical expectation from the contribution to the vacuum energy coming from the quantum field theories is ~ 120 orders of magnitude larger than the observed vacuum energy associated with the acceleration of the Universe. This is the worst theoretical prediction in the history of physics!. It is clear that there is something about the vacuum that we do not understand.

The origin of cosmic acceleration is usually called dark energy, either from a new form of energy or a modification of GR. What dark energy surveys need to answer is if dark energy is distinguishable from a cosmological constant. For that they need to measure the equation of state $w = \frac{p_{\rm DE}}{\rho_{\rm DE}}$, possibly at different redshifts. For a cosmological constant $w = -1$, the so-called ΛCDM model. A further question is if it is possible to distinguish modifications of GR and if it is evolving with redshift. For that it is usually parametrized as [10]:

$$w = w_0 + w_a(1-a) \tag{1.8}$$

Modified gravity predicts a growth rate of gravitational clustering different from GR. Modified gravity theories can be tested measuring both the expansion history and the growth history to see whether they yield consistent results for $H(z)$ or $w(z)$. Nowadays there are no empirically viable modified gravity theories that can explain the observed acceleration, and some theories have already been ruled out [1]. This doesn't mean that one modified gravity theory will not arise in the future.

General relativity is confirmed by a series of observations. Classical tests were done measuring the perihelion precession of Mercury, deflection of light by the Sun, and gravitational redshift of light. Modern tests like gravitational lensing or light time delay also confirm the theory. Besides, the equivalence principle has been tested to extremely high precision by Eötvös torsion balance experiments. The last confirmation of the theory comes from the first direct detection of gravitational waves by LIGO in September 2015. The second observation of gravitational waves was made on December 2015 and announced on 15 June 2016. There are currently six more events (`http://www.ligo.org/news.php`).

1.1.2 The Big Bang Theory

The formulation of the Big Bang theory model started in the 1940s by G.Gamow and its collaborators when they explained the formation and abundance of elements in the early Universe. They realized that the same process that occurs in stars could have happened in the early Universe. This means that the Universe must have been hot and dense enough to allow nucleosynthesis reactions in the early Universe. And then it must have expanded and

cooled to its present state [11, 12]. They also predicted a radiation background, later found in the microwave range (CMB) [13]. This model has some initial condition problems, which can be solved with inflation (sec:1.1.3).

The Big Bang theory explains the evolution of the Universe from the first fraction of a second to our present age, 13.8 billion years later. As we go back in time, the Universe becomes hotter and thus the amount of energy available for particle interactions increases. Interactions go from low energy interactions nowadays (long range gravitational and electromagnetic physics), to atomic physics, nuclear physics, and high energy physics at the electroweak scale. Going further back in time would require assumptions about particle interactions and perhaps about the nature of physical laws themselves.

The Universe must have originated at the Planck era (10^{19} GeV, 10^{-43}s) from a quantum gravity fluctuation. There is not evidence for that and quantum gravity phenomena are still in the realm of physics speculation. However, it is plausible that a primordial era of cosmological inflation originated then. Soon after, the Universe may have reached the Grand Unified Theories (GUT) era (10^{16} GeV,10^{-35}s). At the end of inflation, the energy density of the inflaton field was converted into particles, which soon thermalized and became the origin of the hot Big Bang as we know it, the so-called reheating stage. Since then, the Universe became radiation dominated. It is thought that the matter-antimatter asymmetry originated at baryogenesis (100GeV, 10^{-10}s). Later, baryons (mainly protons and neutrons) formed from their constituent quarks at the quark-gluon phase transition (100MeV,10^{-4}s). When the Universe was at a temperature of 0.8MeV, weak interactions were too slow to keep neutrinos in thermal equilibrium with the plasma, so they decoupled. The first window we have of the early Universe is that of primordial nucleosynthesis (1-0.1MeV,1s-3min), when protons and neutrons were cold enough to form light elements. Much later, matter-radiation equality occurs (1eV,10^5yr). Soon after, atoms are formed (0.3eV,$3 \cdot 10^5$yr), in a process known as recombination. Immediately after, photons decouple from the plasma and travel free since then. Those are the photons observed as the CMB. Finally, the inhomogeneities generated during inflation give rise to structure formation (1-10Gyr) due to gravitational collapse, galaxies and clusters of galaxies are formed. [14].

1.1.3 Inflation

Although the hot Big Bang theory is a robust theory checked with a variety of observations, it has some initial conditions problems. This theory by itself cannot explain the origin of matter and structure in the Universe, why the Universe is so flat, the origin of the primordial density fluctuations that gave rise to all the structures in the Universe or the origin of the Big Bang itself.

In the 1980s a new theory was proposed, simultaneously by Alexei Starobinski [15] in the Soviet Union and Alan Guth [16] in the United States, to solve these problems. The new models of inflation have been developed by Andrei Linde, Andreas Albrecht and Paul Steinhardt [17, 18]. We call inflation to the period of time in the early Universe where the Universe expanded exponentially, due to the constant energy density of a scalar field called the inflaton. This exponential expansion made the space-time stretch even faster than the

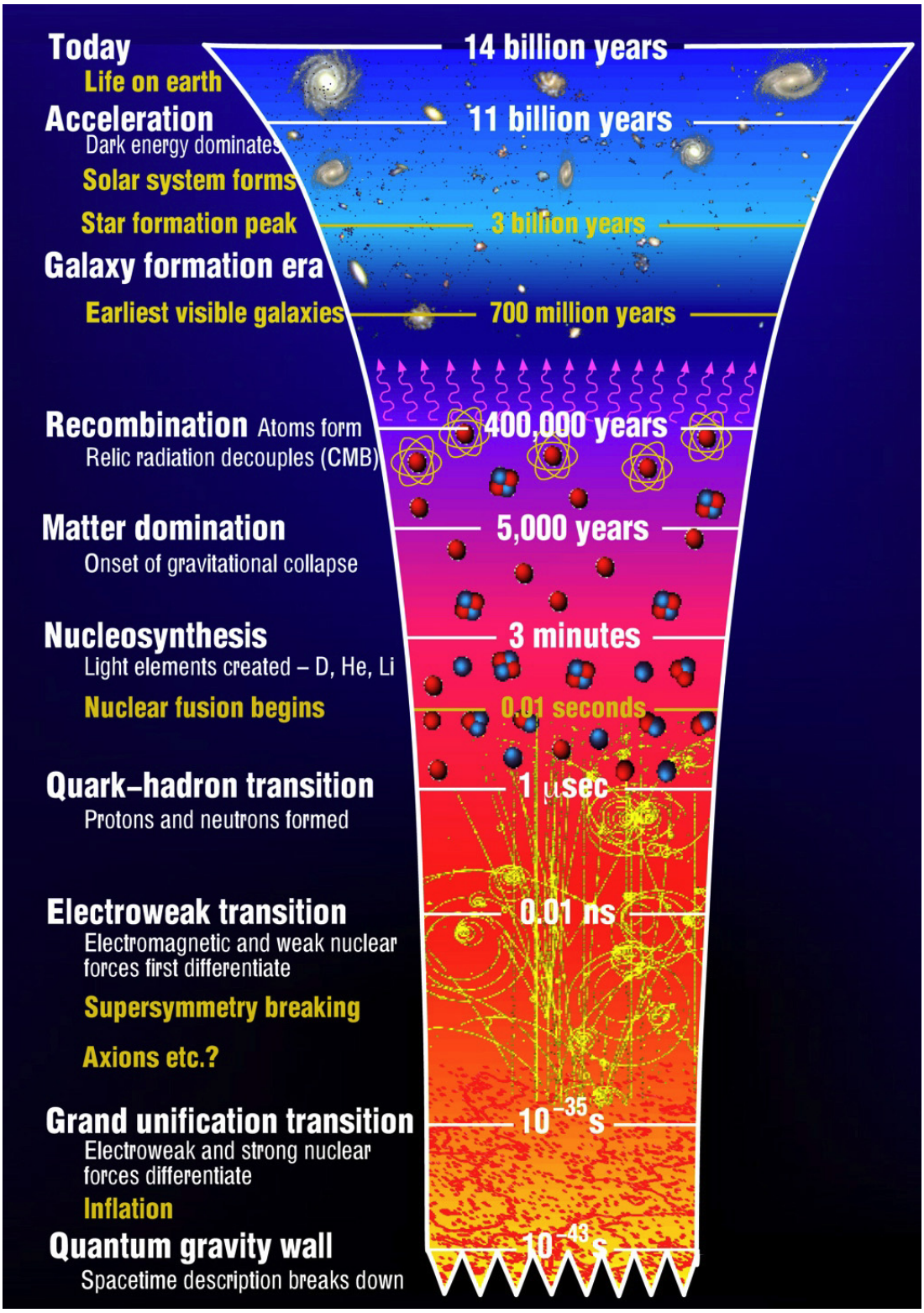


Figure 1.2: Current history of the Universe according to the Standard Model of Cosmology.

speed of light.

This accelerated initial expansion solves many of the problems of the Big Bang theory [14]:

- Flatness Problem: For the Universe to be flat today, $\Omega_K = 0.0008 \pm 0.004$ at 95% confidence level [3], it must have been very flat in the early Universe, at the epoch of nucleosynthesis $|\Omega - 1| < \mathscr{O}10^{-18}$ [19] and at the Planck epoch $|\Omega - 1| < \mathscr{O}10^{-64}$ [20]. This is an extreme fine-tuning initial condition, which can just be fulfilled with special initial values. For other initial values, the Universe would soon collapse, preventing the structure that we see today from being formed.
  Inflation naturally solves this problem, the early exponential expansion of the Universe would lead to the extremely flat initial conditions needed for the Universe to be flat today.

- Homogeneity problem: there are causally connected zones in the Universe, from the early Universe until today the distance a particle could have travelled is given by:

$$d_H(t) = a(t) \int_0^t \frac{dt'}{a(t')} \tag{1.9}$$

  this is called the particle horizon. Why do we see in the CMB that regions which could not have been in causal contact (separated more than 1^o on the sky today) have the same temperature?.
  This question is again answered naturally by inflation, regions which were in causal contact in the early Universe were stretched to non-causal contact distances today.

- Origin of matter and structure in the Universe: the exponential expansion of inflation will dilute any fluctuations in the Universe, but the inflaton, as a quantum filed, will have quantum fluctuations that will induce metric perturbations. The inflaton fluctuations induce waves in the space-time metric that can be decomposed into different wavelengths, all with approximately the same amplitude, that is, corresponding to a scale-invariant spectrum.

The most simple and widespread inflationary theories rely on a single scalar field, the so-called inflaton. The only thing we know about the inflaton is that it is a scalar field ϕ with mass and it has a self-interaction potential $V(\phi)$, but we ignore everything else. We do not know its origin or if it is an effective description of a more fundamental high energy interaction. The Lagrangian of this scalar field is [14]:

$$\mathscr{L}_{\text{inf}} = -\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \tag{1.10}$$

The evolution equation of this homogeneous scalar field in a FRW metric (1.2) is given by

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 \tag{1.11}$$

The dynamics of this fluid can be described as a perfect fluid (1.3) with a time dependent

pressure and energy density given by:

$$\begin{aligned} p &= \frac{1}{2}\dot{\phi}^2 - V(\phi) \\ \rho &= \frac{1}{2}\dot{\phi}^2 + V(\phi) \end{aligned} \tag{1.12}$$

The field evolution equation (1.11) can be written as the energy conservation equation:

$$\dot{\rho} + 3H(\rho + p) = 0 \tag{1.13}$$

If the potential energy density of the scalar field dominates the kinetic energy, $V(\phi) \gg \dot{\phi}^2$, then we see that $p \simeq -\rho$ and then the energy density is constant $\rho \simeq$ cte, so the rate of expansion will be too $H(\phi) \simeq$ cte. This leads to the solution $a(t) \sim \exp(Ht)$, then $\frac{\ddot{a}}{a} > 0$ and we have accelerated expansion.

Nowadays inflation has a very wide variety of models: R^2, chaotic, extended, power-law, hybrid, natural, supernatural, extranatural, eternal, D-term, F-term, brane, oscillating, trace-anomaly driven,... [21]. The simplest inflationary models generate two types of perturbations: density perturbations from the scalar field and gravitational waves from tensor metric fluctuations. Density perturbations are affected by gravitational instability and lead to structure formation, while tensor metric fluctuations can influence CMB anisotropies. A quantity usually used to test inflation is the ratio (r) between the spectra of scalar fluctuations (Δ_R^2) and tensor fluctuations (Δ_t^2):

$$r \equiv \frac{\Delta_t^2}{\Delta_R^2} \tag{1.14}$$

As stated by this theory, quantum fluctuations of the inflaton field left their imprint as tiny perturbations in an otherwise very homogeneous patch of the Universe. Most inflation models predict an approximately scale invariant spectrum $P(k) \propto k^{n_s - 1}$ with $n_s \approx 1$. This has been confirmed by Planck+lensing+BAO+JLA+H_0 68% limits with $n_s = 0.9667 \pm 0.0040$ [3]. The standard cosmology model assumes adiabatic Gaussian perturbations in the initial perturbations. The simplest inflation models predict non-Gaussianities, but too small to be detected by any experiment yet conceived.

1.2 Observational basis of the Standard Model

ΛCDM model is described by a large amount of parameters. These parameters are degenerate and can be reduced to just six parameters (Ω_b, Ω_m, h, n_s, Δ_R^2, τ), taking into account that the value of the radiation density is known with high precision ($T_{CMB} = 2.7255 \pm 0.0006$K [22]) and excluding neutrino energy density. These set of parameters could however be increased if observational evidence for some phenomena like tensor perturbations were found, or spatial curvature was different from zero. These six parameters are the minimum number varied in fits to data, though more parameters are necessary to describe astrophysical processes influencing the data, for example the helium fraction.

Since the beginning of this new century, there has been a huge progress in measurements of the cosmic expansion history and structure growth, leading to much tighter constraints on the parameters of dark energy models. Particularly prominent are measurements of the CMB [3]. Nowadays we can talk about precision cosmology, but parameter precision is the means to an end. The underlying goal of empirical studies of cosmic acceleration is to find out weather the acceleration arises from an extension of GR or from a new energy component. And in the case of a new energy component to find out weather it is constant in space and time or it is a dynamical field.

Different types of observation are sensitive to different subsets of cosmological parameters [2]. An accurate model of the Universe requires a range of different types of observations, with complementary probes providing consistency checks, breaking parameter degeneracies, and setting strong constraints on these parameters. Moreover, the authors in [23] found that overlapping surveys not only provide a better figure of merit (equivalent to 50% larger area) than separate surveys, but they are more robust to systematic errors, such as bias stochasticity or uncertainties in the bias evolution. In the following sections the main probes for dark energy and dark matter and future observations are described.

Most cosmology probes fall in one of these categories:

- Geometrical probes, which probe different weights of the comoving distance:

$$r(z) = \int \frac{dz}{H(z)} \tag{1.15}$$

- Growth factor probes:

$$D''(a) + D'(a)\left[\frac{3}{a} + \frac{H'}{H}\right] = \frac{3}{2}\frac{H_0^2}{H^2}a^{-5}\Omega_m D(a) \tag{1.16}$$

  where the term on the right changes for modified gravity.

1.2.1 Relative abundance of light elements

Big Bang Nucleosynthesis (BBN) predicts the abundances of the light element isotopes D, ^{3}He, ^{4}He, and ^{7}Li [24]. The abundance of light elements depends on the baryon-to photon ratio η_{10}:

$$\eta_{10} = 10^{10}(n_b/n_\gamma) = 274\Omega_b h^2 \tag{1.17}$$

where n_b is the number density of baryons, and n_γ the number density of photons. According to last Planck's results, $\eta_{10} = 6.11 \pm 0.04$ [3]. The observed relative abundances of light elements are in agreement with the predictions of the hot Big Bang theory.

1.2.2 Supernovae Type Ia (SN)

Until 1929 that Hubble discovered that there were galaxies separating from us (all but a few) [7], the Universe was thought to be static. After the discovery of the redshifting of the spectral lines of the galaxies, it was accepted that the Universe was expanding. This expansion was thought to be decelerating due to gravitational attraction of matter. In the late 1990s, two independent supernova surveys provided evidence for the accelerated

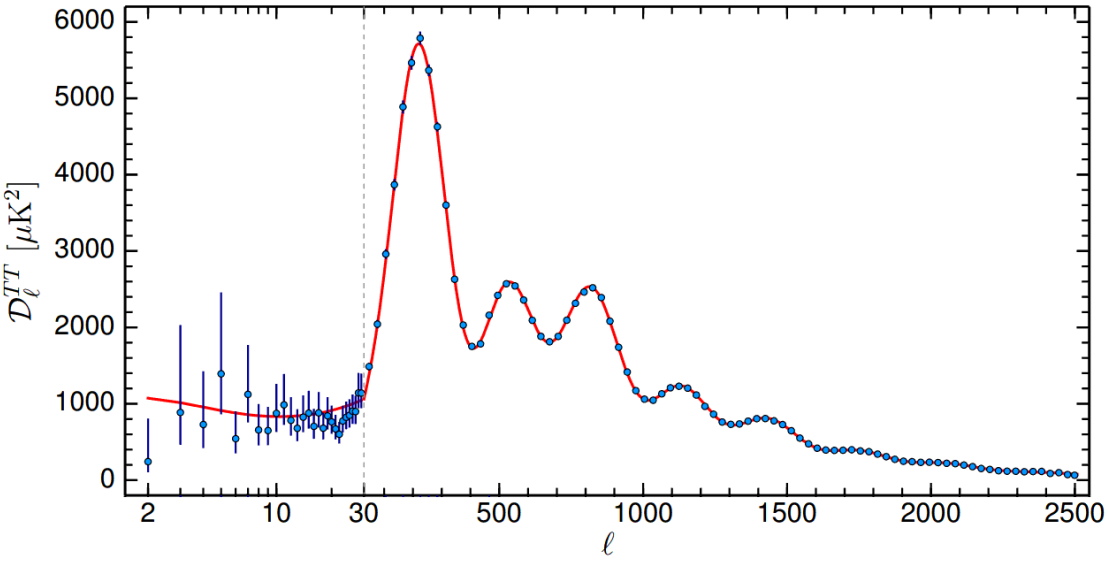


Figure 1.3: Power spectrum of the CMB measured by Planck, at low multipoles ($\ell = 2-49$) plotted on a logarithmic multipole scale and at higher multipoles ($50 < \ell < 2500$) plotted on a linear multipole scale, with the best-fit to the CMB spectrum [26].

expansion, The Supernova Cosmology Project and the High-z Supernova Search Team [8, 9].

Supernovae of type Ia are thermonuclear explosions of a white dwarf in a binary system. The white dwarf gradually accretes mass from a binary companion, until it reaches ~ 1.44 solar masses [25]. Beyond this, they re-ignite and in some cases trigger a supernova explosion. Its peak luminosity can be used as a standard candle, constraining the distance-redshift relation:

$$H_0 d_L(z) = \frac{1+z}{|\Omega_K|^{1/2}} sinn\left[\int_0^z \frac{|\Omega_K|^{1/2} dz'}{\sqrt{(1+z')^2(1+z'\Omega_M - z'(2+z')\Omega_\Lambda}}\right] \tag{1.18}$$

where $sinn(x) = x$ if $K = 0$, $sinn(x) = \sinh(x)$ if $K = -1$ and $sinn(x) = \sin(x)$ if $K = +1$. Although not perfect standard candles, it has been demonstrated that by correcting for a relation between the light curve shape, color, and the luminosity at maximum brightness, the dispersion of the measured luminosities can be greatly reduced. The main systematics here are associated with photometric calibration, dust extinction corrections and redshift evolution of the supernova population [8, 9].

1.2.3 The Cosmic Microwave Background (CMB)

The CMB was discovered in 1965 [27] and it has the most perfect thermal black body spectrum ever measured. The discovery of the CMB is a landmark evidence of the Big Bang origin of the Universe. The CMB is extremely uniform, but it contains tiny anisotropies that are usually expanded in spherical harmonics:

$$\Delta T(\theta,\phi) = \sum_{\ell,m} a_{\ell m} Y_{\ell m}(\theta,\phi) \tag{1.19}$$

The anisotropies are generated by inhomogeneities in the density of the Universe, and they can be expressed as a linear superposition of density perturbations δ_k. As density perturbations are Gaussian random variables with zero mean, so are ΔT and $a_{\ell m}$. Then, we can have a full description of the anisotropies with the two-point correlation function $C(\theta) = \langle \Delta T(\vec{r}_1) \Delta T(\vec{r}_2) \rangle$, where θ is the angle between $\vec{r}_1$ and $\vec{r}_2$. For a homogeneous and isotropic Gaussian random field, the correlation function can be expressed as:

$$C(\theta) = \sum_{\ell m} \sum_{\ell' m'} \langle a_{\ell m} a^*_{\ell' m'} \rangle Y_{\ell m}(\vec{r}_1) Y_{\ell' m'}(\vec{r}_2) \tag{1.20}$$

$$C(\theta) = \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell+1) C_\ell L_\ell(\cos\theta) \tag{1.21}$$

where $\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \delta_{\ell \ell'} \delta_{m m'}$. The angular power spectrum C_ℓ is therefore

$$C_\ell = \langle |a_{\ell m}|^2 \rangle \tag{1.22}$$

Many parameters combine to determine the overall shape of the power spectrum (Figure 1.3), but the location of the peaks probes the spatial geometry, and the relative heights of the peaks probes the baryon density. The mean temperature $T_{CMB} = 2.7255 \pm 0.0006$K [22] can be considered as the monopole component of CMB maps. The largest anisotropy is in the first spherical harmonic, the dipole ($\ell = 1$). This is believed to be caused by the solar system motion relative to the nearly isotropic black body field, as suggested by measurements of radial velocities of local galaxies [28]. Higher multipoles are interpreted as the result of density perturbations in the early Universe.

The first detection of the CMB anisotropies was carried out by the COBE satellite [29]. Later WMAP measured with better precision these anisotropies [30, 31]. The 2015 data release from the Planck satellite [3] has provided the most powerful results to date on the CMB spectrum. The Atacama Cosmology Telescope (ACT) and South Pole Telescope (SPT) are ground-based experiments which extend these results to higher angular resolution, though without full-sky coverage. The CMB shows that the deviations from homogeneity and isotropy at photon decoupling were just a few parts per million. In combination with other data this strongly constrains the spatial geometry, in a manner consistent with spatial flatness and excluding significantly-curved Universes. It also does not show evidence for dynamics of dark energy, being consistent with a cosmological constant.

In principle, the mechanism that produces primordial perturbations could generate scalar, vector and tensor modes. In linear perturbation theory vector modes decay with the expansion of the Universe in the absence of sources. However tensor modes effects might be seen in polarization measurements of the CMB. Polarization can be decomposed in the divergence or E-mode, and the curl or B-mode. The B-mode component can be induced by tensor perturbations, CMB lensing or dust. The determination of a non-zero B-mode signal, subtracting the lensing and dust contributions, would suggest that something like inflation actually occurred. In 2014 BICEP2 reported a B-mode detection from gravitational waves in the early Universe [32]. A later joint analysis with Planck concluded that subtracting the lensing, the signal can be attributed to dust in the Milky Way [33]. In this joint analysis they set a limit on the tensor-to-scalar ratio of $r < 0.12$ at 95% confidence level.

1.2.4 Baryon Acoustic Oscillations (BAO)

Before photon decoupling baryonic matter and radiation propagated as a single fluid. Gravitational attraction of baryons counteracted by radiation pressure created acoustic pressure waves, the so-called Baryon Acoustic Oscillations (BAO). When the Universe was about 380,000 years, photons decoupled and these acoustic waves 'froze'. Photons decoupled from the plasma when their interaction rate could not keep up with the expansion of the Universe and the mean free path became larger than the horizon size: the Universe became transparent. Photons decoupled when the temperature of the Universe fell to 0.26eV, at redshift $z_{\rm dec} \approx 1090$. Soon after this, acoustic waves were 'frozen' at a distance given by the sound horizon at the baryon drag epoch r_s given by:

$$r_s(z_{\rm drag}) = \frac{c}{\sqrt{3}} \int_0^{1/(1+z_{\rm drag})} \frac{da}{a^2 H(a) \sqrt{1+(3\Omega_b/4\Omega_\gamma)}} \text{Mpc/h} \tag{1.23}$$

The traces of the acoustic waves can be seen today in the spatial distribution of galaxies in the Universe, as a bump at the sound horizon scale (r_s) in real space in the two-point correlation function (2PCF) (sec: 2.3), or as a sinusoidal contribution in the Fourier space in the power spectrum ($P(k)$) (sec: 2.2). The sound horizon can be used as a 'standard ruler', since after photon decoupling it is only affected by the expansion of the Universe.

The acoustic phenomenon was first described by Jim Peebles and J.Yu [34], and Rashid Sunyaev and Yakov Zel'dovich [35]. The BAO effects were first seen in the series of peaks in the CMB angular power spectrum [36]. The first significant detection of BAO signal was done in the spatial galaxy distribution in 2005 by Eisenstein et al. [37]. They measured the 2PCF of 46478 LRG's at a medium redshift of $z = 0.35$. The acoustic peak was observed at $r = 100 Mpc/h$, in agreement with the cosmology obtained from the CMB and Supernove of type Ia. Since then several measurements of the BAO have been made at different redshifts. In Figure 1.4 we can see the last BAO measurement of BOSS of the DR12 data release, in real space [38] and in Fourier space [39]. The BAO can also be measured with quasars [40] using the Lyman-α forest. The Lyman-α forest is the absorption features in the spectra of distant quasars due to the presence of neutral hydrogen in the intergalactic medium. Hydrogen at different redshifts absorbs at different observed wavelength the quasar spectrum. The Lyman-α forest is observable in a redshift range inaccessible to galaxies and where theoretical modelling is less dependent on non-linear effects in cosmological structure formation. These factors combine to make Ly-α absorption a promising tracer of mass that is complementary to galaxy tracers.

1.2.5 Weak Lensing (WL)

The light of high redshift galaxies (source) is gravitationally deflected by clustered distributions of matter (lenses) shearing their shape and producing a correlated pattern of apparent ellipticities [41, 42]. The size of the deflection angle depends both on the mass of the foreground clustered distribution and upon the ratios of distances between observer, lens, and source. By studying this effect statistically, assuming that galaxies are oriented randomly in the absence of lensing, one can infer the mass distribution in the foreground of these source galaxies (Figure 1.5). As light deflection is affected by gravity, it is a powerful tool to measure the mass distribution in the Universe, since the gravitational potential is

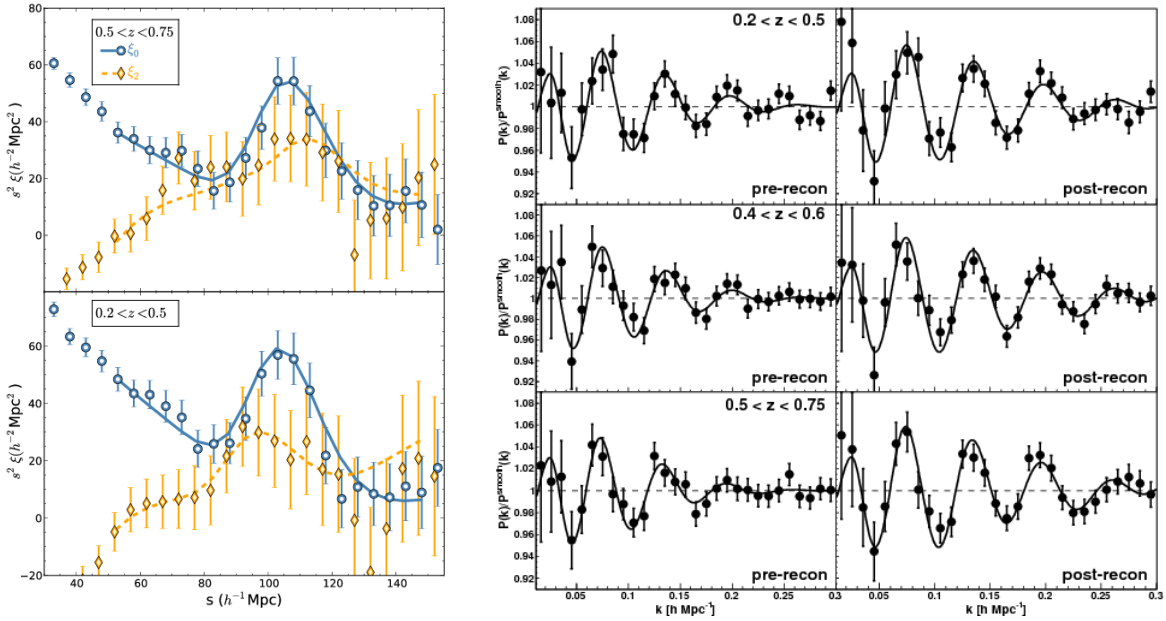


Figure 1.4: The DR12 BAO measurements from BOSS along with their fits. The 2PCF post-reconstruction in the left panel [38] and the Power spectrum [39] in the right panel.

affected by both baryonic and dark matter. Gravitational lensing observations probe the dark energy via both the expansion history and the growth history of density fluctuations. The completed CFHTLS survey (stage-II) measured shear in the 155 sq. deg. wide field to $i < 24.5$ [43, 44]. The last cosmological constraints from weak lensing are those from KIDS survey [45].

1.2.6 Clusters of Galaxies (CL)

A cluster of galaxies is a large amount of galaxies held together by their gravitational attraction. The largest ones are around 10^{15} solar masses, and they are the largest gravitationally-collapsed structures in the Universe. Clusters of galaxies probe structure growth by constraining $\sigma_8\Omega_m^{\alpha}$, where $\alpha \approx 0.3-0.5$. These halos can be identified as dense concentrations of galaxies or through the signatures of hot gas in X-ray emission or Sunyaev-Zel'dovich distortion of the CMB [46]. The critical challenge in cluster cosmology is calibrating the relation $P(M_{halo}|O)$ between the halo mass and the observable used for cluster identification.

1.2.7 Redshift Space Distortions (RSD)

The distance to the galaxies is estimated using the redshift (given a cosmological model), measured from the shift of the spectral lines in their emission spectra caused by the expansion of the universe. However, due to the presence of density perturbations, galaxies have non-zero radial peculiar velocities, which alter their measured redshift through an additional Doppler effect. Peculiar velocities v are defined as deviations from the Hubble

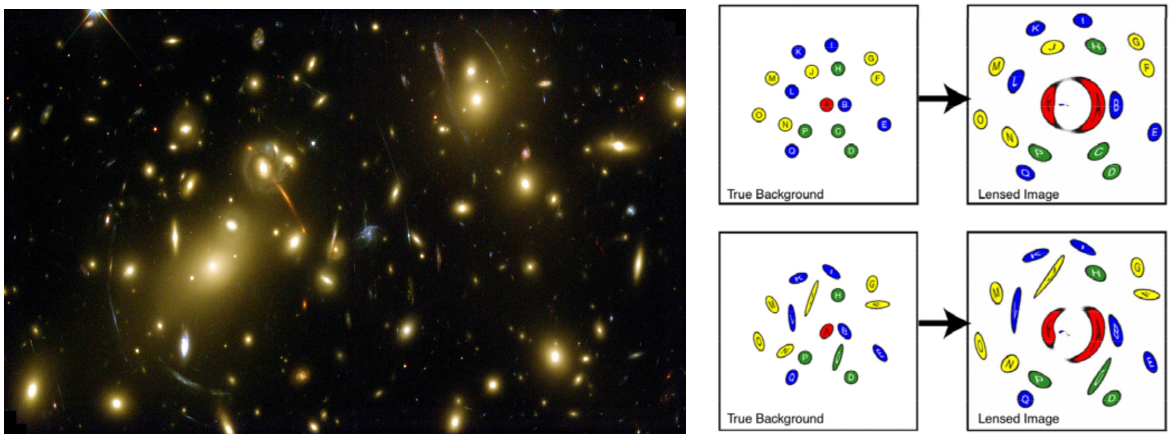


Figure 1.5: At left is an image of a galaxy cluster from the Hubble Space Telescope, exhibiting arc-like images of faint background galaxies that are characteristic of strong gravitational lensing. At right: the upper panels show a fictitious collection of circular background galaxies before (left) and after (right) lensing by a foreground mass concentration. While galaxy A, on axis, is grossly distorted into a ring, all the other galaxies undergo a slight shearing by the lens. On the lower panels, the galaxies have a variety of initial shapes, so the lensing shear pattern is less obvious, but would be detectable by statistical analysis. Image from [2].

flow due to local inhomogeneities:

$$\dot{r} = (\dot{ax}) = \dot{a}x + a\dot{x} = Hr + v \tag{1.24}$$

where $r = ax$ is the proper distance and x is the comoving distance. At large scales we will have distortions in the density fluctuations due to peculiar velocities of galaxies. The structures we will see in the large scale distribution of galaxies will be denser and the voids emptier, i.e., a compression of the correlation function in the line of sight known as the Kaiser effect [47]. At small scales, random velocities of galaxies inside clusters produce a radial elongation in the galaxy distribution pointing to the observer, known as the fingers of God (FOG) [48]. These effects are shown in Figure 1.6.

Redshift-Space Distortions (RSD) probe structure growth by constraining the parameter combination $f(z)\sigma_8(z)$, where

$$f(z) = \frac{d\ln\delta(a)}{d\ln a} \tag{1.25}$$

is the growth rate, and $\delta(a)$ the density contrast. $f(z)$ is usually parametrized as $f(z) \approx \Omega_m^{\gamma}(z)$, where γ is the gravitational growth index. For the General Relativity case $\gamma = 0.55$ [49]. This method is limited by uncertainties in theoretical modelling of non-linear gravitational evolution and non-linear bias between galaxy and matter distributions, which limits its application to large scales. One of the last measurements is the one from SDSS [50, 51].

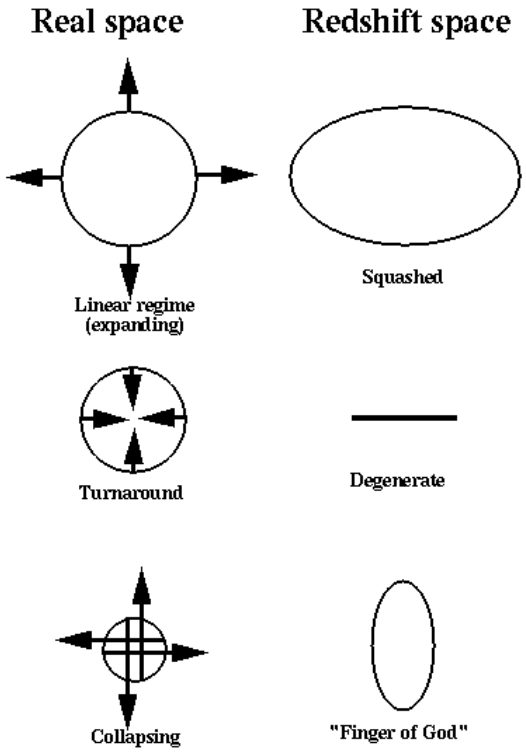


Figure 1.6: Schematic representation of how real-space structures (left column) look in redshift-space (right column).

1.2.8 Dark Matter

Dark Matter (DM) is a non-luminous and non-absorbing matter which does not emit or interact with electromagnetic radiation, such as light, and is thus invisible to the entire electromagnetic spectrum [52]. Although dark matter has not been directly observed, its existence and properties are inferred from its gravitational effects such as the motions of visible matter, gravitational lensing, its influence on the Universe's large-scale structure, on galaxies, and its effects in the CMB. Whether it is relativistic (hot) or non-relativistic (cold) can be inferred from observations. Analyses of structure formation in the Universe indicate that most DM should be 'cold', since relativistic particles tend to diffuse from one concentration of matter to another, thus preventing the growth of structure on small scales.

The current most accurate determination of Ω_m comes from global fits of cosmological parameters to a variety of observations [3]. Although the existence of dark matter is well established there are several dark matter candidates, and not direct or indirect confirmation has yet been found for any of them. Candidates for non-baryonic DM must be stable on cosmological scales, they must interact weakly with electromagnetic radiation, and they must have the right relic density. Primordial black holes [53], axions, sterile neutrinos and weakly interactive massive particles (WIMPS) accomplish these requirements.

Direct detection experiments aim to observe low-energy recoils (typically a few keVs) of nuclei induced by interactions with dark matter particles, which (in theory) are passing through the Earth. Examples of underground laboratories which house direct detection

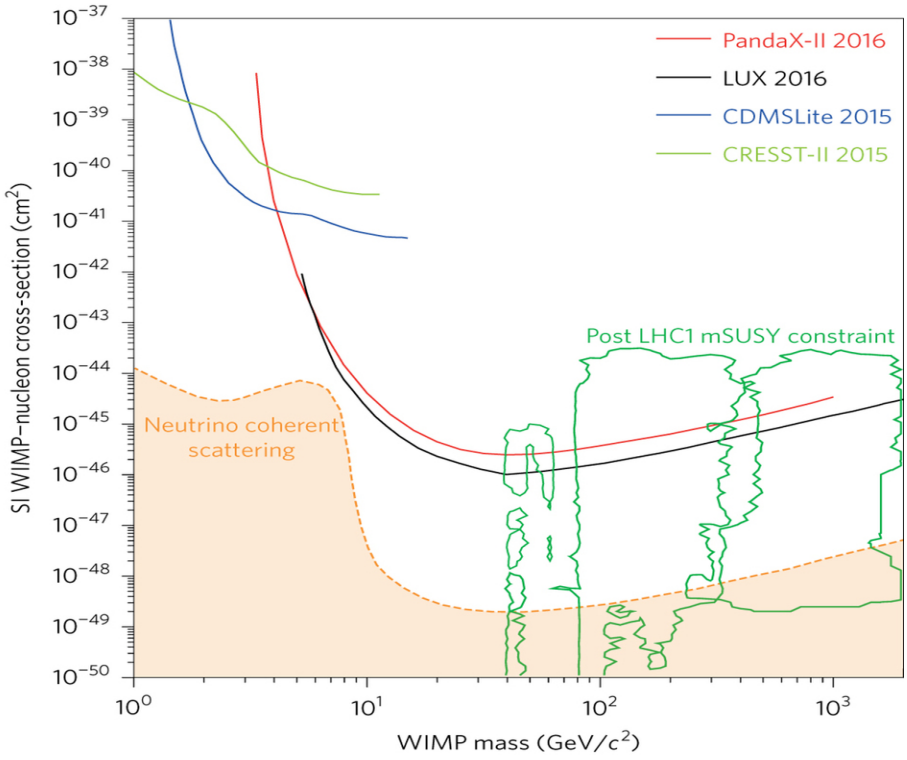


Figure 1.7: WIMP cross sections from PandaX-II, LUX, SuperCDMS (CDMSLite) and CRESST-II. The neutrino coherent scattering background curve data and the post-LHC-Run1 minimal-SUSY model allowed contours are also shown [54].

experiments include the Stawell mine, the Soudan mine, the SNOLAB underground laboratory at Sudbury, the Gran Sasso National Laboratory, the Canfranc Underground Laboratory, the Boulby Underground Laboratory, the Deep Underground Science and Engineering Laboratory and the China Jinping Underground Laboratory. Currently there has been no well-established claim of dark matter detection from a direct detection experiment, leading instead to strong upper limits on the mass and interaction cross section with nucleons of such dark matter particles that can be seen in Figure 1.7 [54].

Indirect detection experiments search for the products of the self-annihilation or decay of dark matter particles. These processes could be detected indirectly through an excess of gamma rays, antiprotons or positrons emanating from high density regions in our galaxy or others. A major difficulty inherent in such searches is that there are various astrophysical sources which can mimic the signal expected from dark matter. These methods are complementary to direct detection. Currently there are several neutrino telescopes such as MACRO, BAKSAN, SuperKamiokande, Baikal AMANDA, ANTARES, NESTOR, IceCube and Pierre Auger Observatory.

1.3 Galaxy Surveys

In a galaxy redshift survey galaxies are located in space with three observable quantities: angular coordinates, right ascension (RA, α) and declination (dec, δ) in equatorial coordinates, and its redshift z. From the redshift we can just estimate the actual distance to the objects assuming a particular cosmological model.

There are two different approaches to measure z:

- Spectroscopic redshifts. In spectroscopic surveys the full spectrum of each galaxy is measured and used to determine its redshift with a very good precision. From the photometry galaxies are selected and located in the sky and the spectrum of every galaxy is measured with a fibre. This process is very time-consuming, and can only be done for a subset of the observed sample. Surveys such as 2dFGRS [55], 6dF [56], WiggleZ [57] or SDSS [58].

- Photometric redshifts. The redshift is estimated from photometry in a few broadband filters. There are two main approaches for measuring photometric redshifts:
  * template fitting methods based on the measured broadband galaxy spectral energy distribution (SED) (e.g. Hyperz, BPZ, LePhare, EAZY).
  * training methods based on machine-learning algorithms (e.g. ANNz, ArborZ, TPZ).

  The resulting photometric redshift (photo-z here on) has a larger uncertainty, σ_z, than a spectroscopic redshift. This error can usually be parametrized as a function of redshift as $\sigma_z = \sigma_0(1+z)$ due to the stretching of the spectra with redshift for a filter set of constant resolution. The magnitude of this error depends on the filter set, the type of galaxy and the method used. For most of the surveys listed below, it is not expected to be below $\sigma_0 \sim 0.03$. Due to this large error, structures are washed off along the line of sight. The Dark Energy Survey (DES) [59] and PanSTARRS [60] are two ongoing large area photometric redshift surveys, and LSST [61] will start in the future.

Most spectroscopic instruments can just measure one spectra per fiber, having to select previously the targets to be observed (usually using imaging). The are also instruments capable of taking the spectra of all the objects in a given field (IFUs), the problem is that the exposure times are longer and the technology is still no widely used. On the other hand, photometric surveys can estimate redshifts for all the observed sample (with better or worse accuracy), and therefore the shot noise problem that Spectroscopic redshifts may have at high redshifts is largely alleviated. The big uncertainty in the photo-z prevents a correct reconstruction of the radial information, and only allows in practice a study of the angular statistics of galaxy clustering, such as the angular 2PCF $w(\theta)$ or the angular power spectrum C_l.

Two Spanish projects, PAU [62] and JPAS [63] will aim to obtain photometric redshifts using medium-band filters, an intermediate approach between spectroscopy and photometry, which will improve the photo-z accuracy significantly. Through 40 narrow-band filters, the PAU survey will achieve a high accuracy photo-z (0.3%) for $i_{AB} \sim 23$ [64].

Besides redshifts and angular positions, measuring the photometry of the galaxies is a crucial task. The colours, or the relative magnitudes in a given band, contain information about the galaxy type, and this information may be used to detect sources of systematic uncertainties and to minimize their effect, as occurs for example with the galaxy clustering bias. Other galaxy properties may be used as observational probes for cosmology (e.g. galaxies' shapes and sizes for Weak Lensing).

1.4 Current status and future Observations

Figure 1.8 [65] shows the distance-redshift relationship from measurements of different observations and surveys. They are all in good agreement with a flat ΛCDM model. In Figure 1.9 [1] the constraints on the present matter fraction Ω_m and dark energy model parameters are shown. While CMB, BAO, and SN data sets considered are mutually consistent with a flat ΛCDM model, tensions arise with other cosmological measurements, like direct measurements of H_0 from Hubble Space Telescope (HST) (Figure 1.10) [1]. There is also a 2σ tension between the $\sigma_8 - \Omega_m$ combination predicted by Planck-WP+CMB+BAO for ΛCDM and the lower value implied by the weak lensing measurements (Figure 1.11) [1]. The constraints from recent cluster analyses are not in perfect agreement either. Nonetheless, on balance the cluster analyses, like the weak lensing analyses, favour lower $\sigma_8\Omega_m^\alpha$ than the value extrapolated forward from Planck+WP assuming flat ΛCDM. Redshift-space distortion analyses also tend to favour lower values, though statistical errors are still fairly large. Going from ΛCDM to wCDM (a cosmological constant $w_0 \neq -1$) does not solve the problem, this highlights the importance of understanding and reducing systematic effects. This tension might be solved in the future with larger samples, better data and better modelling, or it can finally reflect a deviation from GR. In this sense, combining different probes and surveys is essential to break degeneracies between the different parameters. The authors from [66] present cosmological constraints from DES using a combined analysis of angular clustering and weak gravitational lensing. They do also a joint analysis with Planck CMB data, BAO and SNIa measurements and improve the constraints on σ_8 and w from just Planck alone. This is in part because DES provides modest constraints on H_0 which help break degeneraciey between h and Ω_m in the CMB. Also Planck dataset prefers higher values of σ_8 and h than DES data. Other possible physical resolution for the tensions observed in Figure 1.11, could come from dark energy models with significant time evolution, from a massive neutrino component that suppresses low redshift structure growth, or from decaying dark matter that reduces Ω_m at low z.

Dark energy is one of the most important discoveries in cosmology, with profound implications for astronomy, high-energy theory and general relativity. The dark energy program is focused on [2]:

- Determine whether the accelerating expansion is consistent with a cosmological constant.
- Measure any time evolution of the dark energy.
- Search for a possible failure of GR.

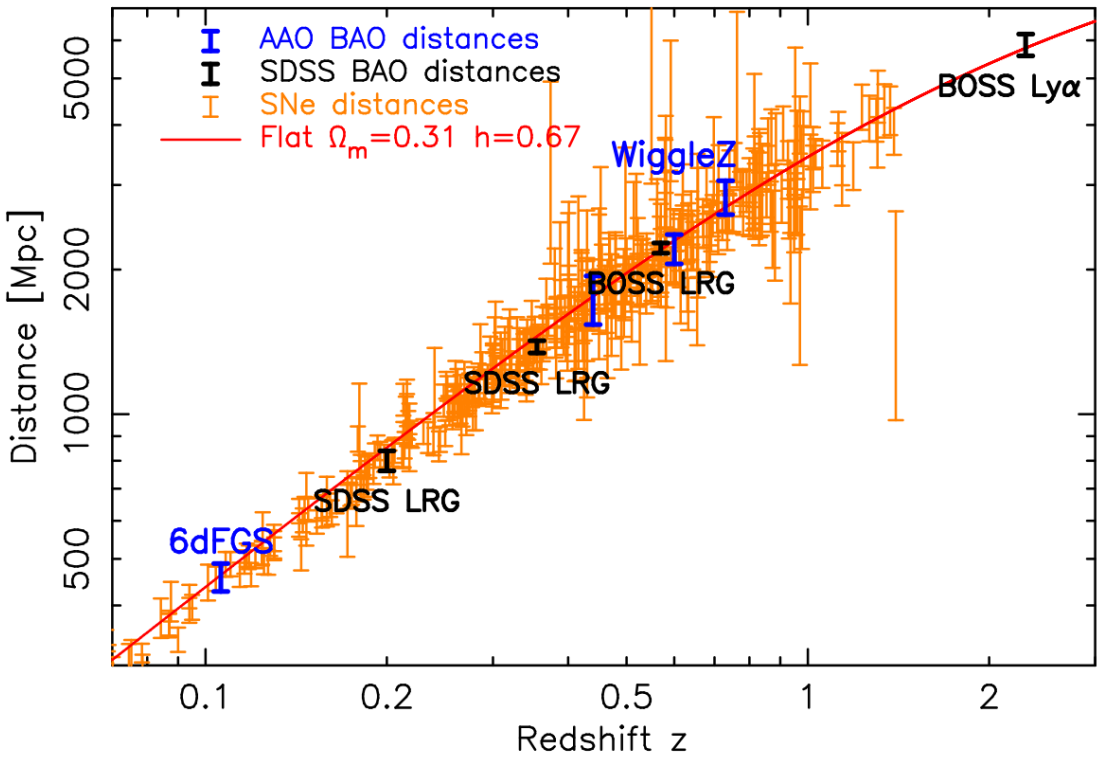


Figure 1.8: The cosmic distance scale with redshift. This modern version of the 'Hubble Diagram' combines data from SN Ia and BAO in the LRG SSDS, BOSS, 6dFGRS, and WiggleZ galaxy surveys and from the BOSS Lyman-alpha at high redshift. For the purposes of this Figure, BAO measurements of $D_L(z)$ have been converted into $D_V(z)$ assuming a Hubble parameter $H(z)$ for a flat ΛCDM model with $\Omega_m = 0.29$ and $h = 0.69$, indicated by the solid line in the Figure, and SNe measurements of $D_L(z)$ have been plotted assuming $D_A(z) = D_L(z)/(1+z)^2$ [65, 67].

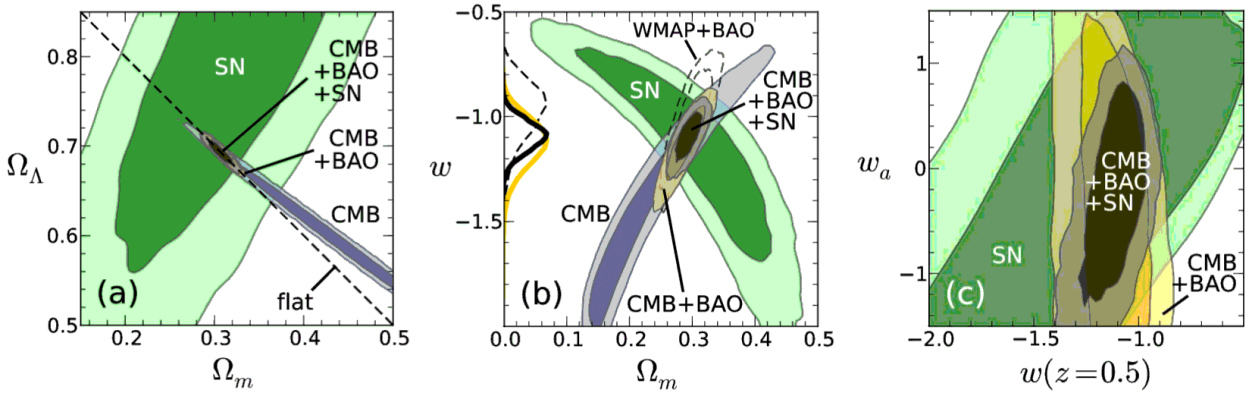


Figure 1.9: Constraints on the present matter fraction Ω_m and dark energy model parameters. Dark and light shaded regions indicate 68.3% and 95.4% confidence levels, respectively. 'CMB' is Planck+WP, 'BAO' is the combination of SDSS-II, BOSS, and 6dFGS, and 'SN' is Union2. In the left panel the present dark energy fraction Ω_Λ vs. Ω_m, assuming a ΛCDM model. In the central panel the dark energy equation of state w vs. Ω_m, assuming a constant value of w. The dashed contours show the 68.3% and 95.4% CL regions for the combination of WMAP9 and BAO data. Curves on the left vertical axis show the probability distributions for w (normalized arbitrarily), after marginalizing over Ω_m, for the CMB+BAO and CMB+BAO+SN combinations (yellow and black, respectively), using Planck+WP CMB data, and for the WMAP9+BAO combination (dashed black). Right panel shows the constraints on the two parameters of the dark energy model with a time-dependent equation of state. Figure from [1].

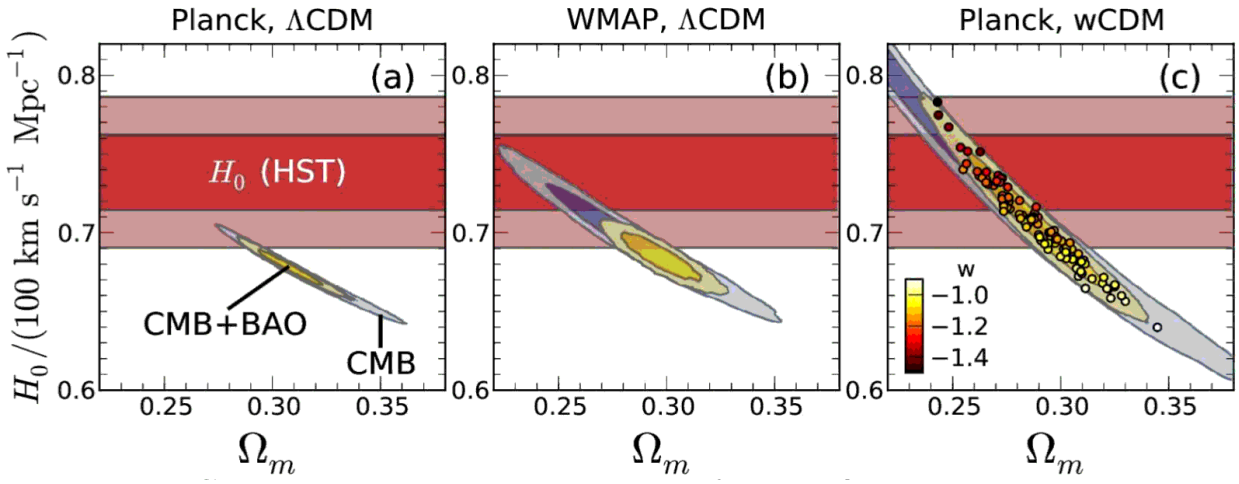


Figure 1.10: Constraints on the present matter fraction Ω_m and the Hubble constant H_0 from various combinations of data, assuming flat ΛCDM (left and middle panels) or a constant dark energy equation of state w (right panel). Dark and light shaded regions indicate 68.3% and 95.4% confidence levels, respectively. The right panel also shows 100 Monte Carlo samples from the CMB+BAO constraints with the value of w indicated by the colors of the dots. 'CMB' is Planck+WP in the outer panels and WMAP9 in the middle panel, 'BAO' is the combination of SDSS-II, BOSS, and 6dFGS, and 'H_0 (HST)' is the HST constraint from [68]. Figure from [1].

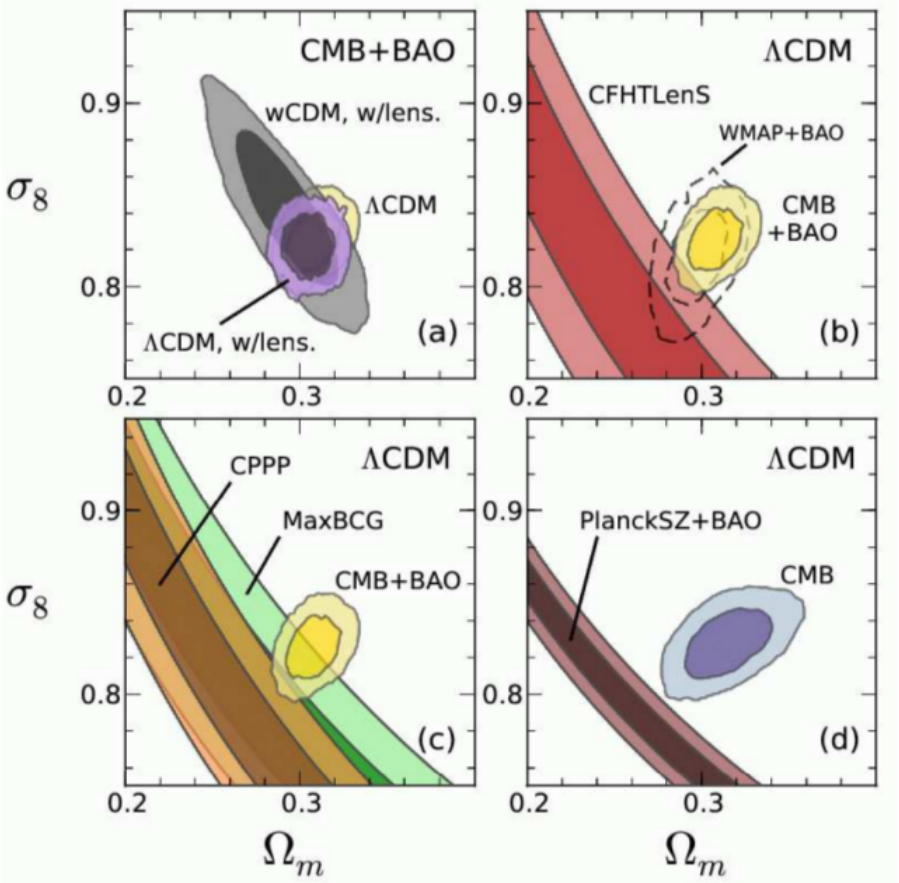
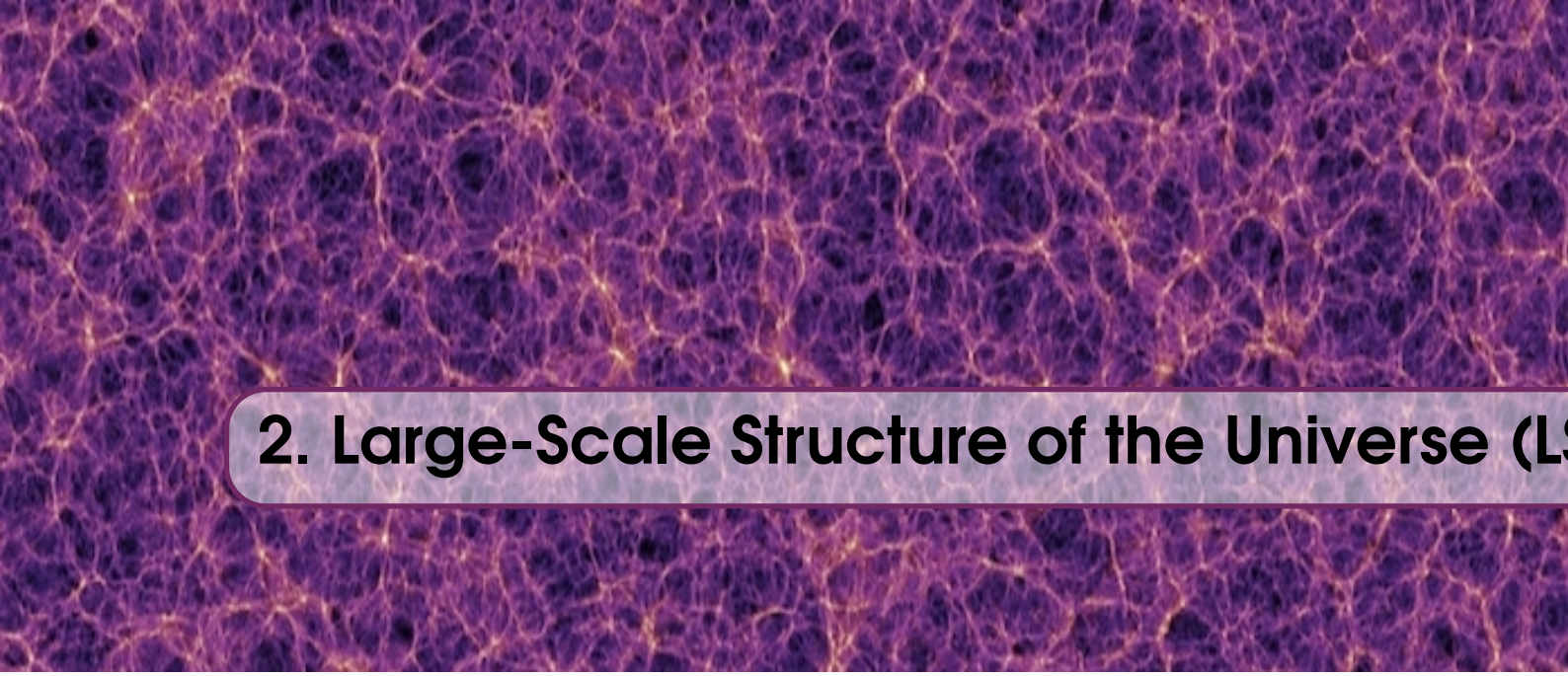


Figure 1.11: Constraints on the present matter fraction Ω_m and the present matter fluctuation amplitude σ_8. Dark and light shaded regions indicate 68.3% and 95.4% confidence levels, respectively. The upper left panel compares CMB+BAO constraints (using the same data sets as in Fig. 1.9) for ΛCDM with and without CMB lensing, and for a constant w model (including CMB lensing). The other three panels compare flat ΛCDM constraints between various dark energy probes, including weak lensing (upper right panel) and clusters (lower panels). Figure from [1].

The Dark Energy Task Force joined in 2005 to study and establish a joint dark energy program [2]. They strongly recommended an aggressive program to explore dark energy combining different techniques with different strengths and sensitivity in different ways to the dark energy properties and to other cosmological parameters. They found that the most powerful probes to constrain dark energy are: BAO, CL, SN and WL, and classified the dark energy program in stages:

- Stage I: What was known by 2006 (Type Ia Supernovae, CMB anisotropies, Weak Lensing and BAO).
- Stage II: Completion of ongoing projects by 2006 (Canada-France-Hawaii Telescope Supernova Legacy Survey (CFHT-SNLS), ESSENCE, Sloan Digital Sky Survey-II (SDSS-II), Center for Astrophysics Supernova Program, Nearby Supernova Factory, Katzman Automatic Imaging Telescope (KAIT), Carnegie Supernova Project (CSP), QUEST Survey, HST Searches for High Redshift Supernovae, PanSTARRS-1, Parallel Imager for Southern Cosmological Observations (PISCO), South Pole Telescope (SPT), Atacama Cosmology Telescope (ACT), XMM Cluster Survey (XCS), Red-Sequence Cluster Survey 2 (RCS2), Deep Lensing Survey (DLS), Kilo-Degree Survey (KIDS) and DEEP2 Galaxy Redshift Survey).
- Stage III: near-term, medium-cost, projects proposed in 2006 (Dark Energy Survey (DES), Hobby-Eberly Telescope Dark Energy Experiment (HETDEX), Wide-Field Multi-Object Spectrograph (WFMOS), Pan-STARRS-4, One-Degree Imager (ODI), One Thousand Points of Light Spectrograph, ALPACA, Cluster Imaging eXperiment (CIX), Cornell-Caltech Atacama Telescope (CCAT) and Physics of the Accelerating Universe (PAU)).
- Stage IV: Future big projects (EUCLID, Large Synoptic Survey Telescope (LSST), Joint Dark Energy Mission (JDEM: Dark Energy Space Telescope (DESTINY), Joint Efficient Dark-energy Investigation (JEDI) and Supernova Acceleration Probe (SNAP)), Square Kilometer Array (SKA),10K X-Ray Cluster Survey, Constellation-X, The Giant Segmented Mirror Telescope (GSMT) and James Webb Space Telescope (JWST)).

Nowadays we are finishing the Stage III era, and will enter in some years the Stage IV. Projects like the Dark Energy Survey are now finishing their observations and will publish their final results in a few years. LSST in USA (ground-based 8.4m telescope and a 3 Gigapixel camera, [69, 70]) and Euclid (space project, [71, 72]) in Europe are the two main big projects for the future. LSST will begin taking data by 2021, and Euclid spacecraft is planned to be launched for 2020. Another spectroscopic survey is DESI [73] which is a merge of the previous BigBoss [74] and DESpec collaborations [75]. DESI is expected to start in 2018.



2. Large-Scale Structure of the Universe (LS

Galaxy surveys are an important tool for Cosmology. Understanding the formation and distribution of Large Scale Structures (LSS) in the Universe can give us a lot of information. In the Standard Model of cosmology, the structure is formed from the initial perturbations created at the end of Inflation. In this chapter I describe how density fluctuation evolution can be described with perturbation theory (sec: 2.1) and the statistical tools used to describe it (sec: 2.2, 2.3, 2.4). Due to non-gravitational processes this linear perturbation theory fails and higher orders must be taken into account (sec: 2.5, 2.6). I also describe CiC as a simple but powerful tool to study LSS (sec: 2.7). Finally, I describe how the spatial distribution of galaxies is related to that of the underlying mass (sec: 2.8).

2.1 Linear Perturbation Theory (LPT)

Galaxy formation in recent epochs is well described by Newtonian gravity and magnetohydrodynamics, but a relativistic treatment is needed for perturbations on scales larger than the horizon before horizon crossing. The horizon size is the coordinate distance travelled by a photon since the beginning of the universe, $d_H \sim H^1$, i.e. the size of causally connected regions in the universe. The relevant components of the Universe are dark matter, baryons, photons and neutrinos. The cold dark matter (CDM) and the baryon components behave like collisionless and collisional fluids, respectively, while photons and neutrinos require a phase-space description governed by the Boltzmann transport equation. Here we are interested in the evolution of the cosmic fields during the matter-domination epoch. In this case, diffusion effects are negligible and the evolution of fluctuations can be described in terms of a perfect fluid (with density ρ, pressure p, and velocity u), and the equations that describe conservation of mass and momentum (the continuity, Euler and Poisson equations) [49]:

$$\left(\frac{\partial \rho}{\partial t}\right)_{\vec{r}} + \nabla_{\vec{r}} \cdot \rho \vec{u} = 0 \tag{2.1}$$

$$\rho \left[\left(\frac{\partial \vec{u}}{\partial t}\right)_{\vec{r}} + (\vec{u} \cdot \nabla_{\vec{r}})\vec{u} \right] = -\nabla_{\vec{r}} p - \rho \nabla_{\vec{r}} \Phi \tag{2.2}$$

$$\nabla^2_{\vec{r}} \Phi = 4\pi G \rho \tag{2.3}$$

We can separate the density into the mean $\langle\rho\rangle \equiv \langle\rho\rangle_{\vec{x}}$ and its perturbation δ:

$$\rho(\vec{x},t) = \langle\rho\rangle(t)(1+\delta(\vec{x},t)) \tag{2.4}$$

and then define the density contrast $\delta(\vec{x},t)$ as:

$$\delta(\vec{x},t) = \frac{\rho(\vec{x},t)}{\langle\rho\rangle(t)} - 1 \tag{2.5}$$

We can write the equations in terms of the peculiar velocity v:

$$\vec{u} = \dot{a}\vec{x} + a\dot{\vec{x}} = H\vec{r} + \vec{v}(\vec{r}/a,t) \tag{2.6}$$

If we change to comoving coordinates $\vec{r} = a(t)\vec{x}$, and separate the potential into the homogeneous part Φ and its perturbation ϕ, we can combine the previous equations:

$$\frac{\partial\vec{v}}{\partial t} + \frac{1}{a}(\vec{v}\cdot\nabla)\vec{v} + \frac{\dot{a}}{a}\vec{v} = -\frac{1}{\rho a}\nabla p - \frac{1}{a}\nabla\phi \tag{2.7}$$

$$\dot{\delta} + \frac{1}{a}\nabla\cdot(1+\delta)\vec{v} = 0 \tag{2.8}$$

$$\nabla^2\phi = 4\pi G\rho\delta \tag{2.9}$$

Linear Perturbation Theory (LPT) can be applied when the amplitude of the density perturbations is small, $\delta \ll 1$. Then, equations 2.7 and 2.8 can be combined to give:

$$\frac{\partial^2\delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial\delta}{\partial t} = \frac{\nabla^2 p}{\rho a^2} + 4\pi G\rho\delta \tag{2.10}$$

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot\vec{v} = 0 \tag{2.11}$$

These equations describe density perturbations in an expanding Universe. The solution can be separated in a growing mode, $D_1 f_1$, and a decaying mode, $D_2 f_2$:

$$\delta(\vec{x},t) = f_1(\vec{x})D_1(t) + f_2(\vec{x})D_2(t) \tag{2.12}$$

The growth of structure is built from the growing mode [49]:

$$D(a) = \frac{5}{2}\Omega_M H_0^2 H(a)\int_0^a \frac{da'}{(a'H(a'))^3} \tag{2.13}$$

This equation is usually normalized to $D(a=1)=1$:

$$D_+(a) = \frac{D(a)}{D(1)} \tag{2.14}$$

and $D_+(a)$ is the so-called growth factor.

Then, when matter perturbations are small, they grow according to linear theory:

$$\delta(\vec{x},t) = \delta(\vec{x},t_0)\frac{D_+(t)}{D_+(t_0)} \tag{2.15}$$

2.2 Power Spectrum $P(k)$

We can now change to Fourier space and write the density contrast as independent modes:

$$\delta(\vec{x},t) = \int d^3\vec{k}\delta_k(t)e^{i\vec{k}\cdot\vec{x}} \tag{2.16}$$

The power spectrum is defined as:

$$\langle|\widehat{\delta}(\vec{k})\widehat{\delta}^*(\vec{k'})|\rangle = (2\pi)^3\delta^D(\vec{k}-\vec{k'})P(\vec{k}) \tag{2.17}$$

where δ^D is the Dirac delta.

If the distribution is Gaussian, the power spectrum gives us all the statistical information. Even higher order moments would be proportional to the second order moment, and odd moments would be zero. In the standard model, in most simple inflationary models initial fluctuations are Gaussian. An important property of this type of fluctuations is that the Fourier transform of the field is also Gaussian.

The power spectrum is the Fourier transform of the two-point correlation function $\xi(\vec{r})$ (sec:2.3):

$$P(\vec{k}) = \int dr^3\xi(\vec{r})e^{-i\vec{k}\cdot\vec{r}} \tag{2.18}$$

$$\xi(\vec{r}) = \frac{1}{(2\pi)^3}\int dk^3P(\vec{k})e^{i\vec{k}\cdot\vec{r}} \tag{2.19}$$

For a homogeneous and isotropic distribution field, $\xi(\vec{r}) = \xi(r)$ and $P(\vec{k}) = P(k)$, and the previous equations can be simplified further:

$$P(k) = 4\pi\int_0^\infty \xi(r)\frac{sin(kr)}{kr}r^2dr \tag{2.20}$$

$$\xi(r) = \frac{1}{2\pi^2}\int_0^\infty P(k)\frac{sin(kr)}{kr}k^2dk \tag{2.21}$$

The primordial matter power spectrum is a power law, $P_M(k) \propto k^{n_s}$, with $n_s = 0.9667 \pm 0.0040$ [3]. In linear theory, Fourier modes evolve independently, $\delta_k(t) \propto D_+(t)$. Then the power spectrum is preserved, but its amplitude grows as $D_+^2(t)$:

$$P_M(k,z) = D_+^2(z)P_L(k) \tag{2.22}$$

where $P_L(k) = k^{n_s}T^2(k)$ is the linear power spectrum that derives from the linear evolution of density fluctuations through the radiation domination era and the resulting decoupling of matter from radiation. $T(k)$ is the transfer function that describes the evolution of density perturbations through decoupling ($T(0) \equiv 1$). This evolution can be followed using general relativistic Boltzmann numerical codes like CAMB [76], CMBFAST [77] or CLASS [78]. The transfer function links the amplitude of primordial perturbations with the amplitude of perturbations at some time after matter and radiation equality $a_{\text{late}} \gg a_{\text{eq}}$. And the growth factor $D_+(a)$ relates the amplitude of the perturbation at this late time a_{late} with its value at any later time $a \gg a_{\text{eq}}$. Since scales entering the horizon before and after equality have

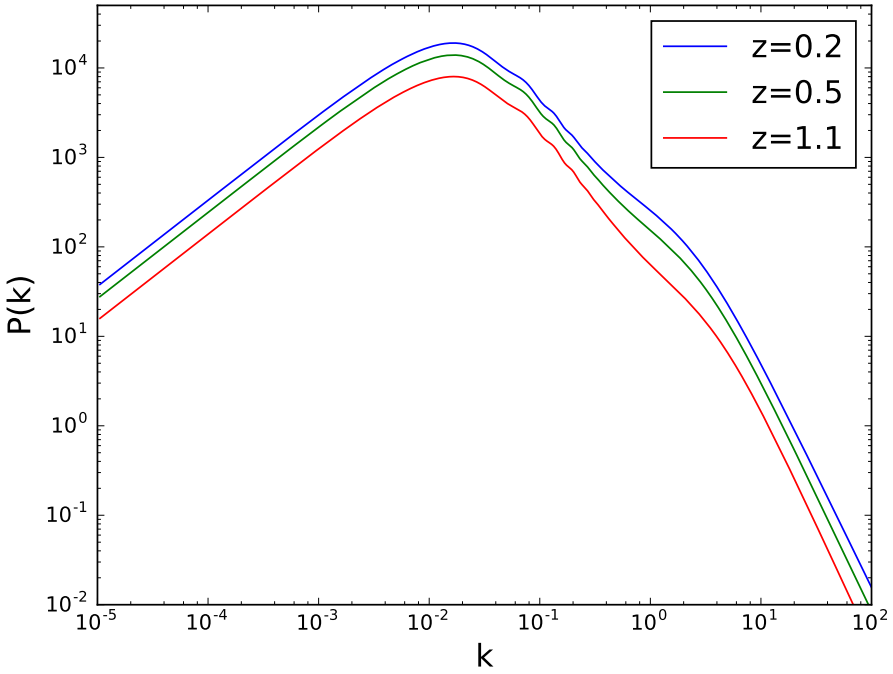


Figure 2.1: Matter power spectrum obtained with CAMB using Planck Cosmology [3] for different redshifts.

very different histories, we must expect a feature of some sort in the power spectrum of matter perturbations at the scale of the horizon at equality k_{eq}:

$$T_k \propto \begin{cases} 1, & k \ll k_{\text{eq}} \\ k^{-2}\ln(k), & k \gg k_{\text{eq}} \end{cases} \tag{2.23}$$

In Figure 2.1 we can see the matter power spectrum obtained with CAMB using Planck Cosmology [3] for different redshifts. We can see the different behaviour after and before the equality scale k_{eq}:

$$P(k) \propto \begin{cases} k & k \ll k_{\text{eq}} \\ k^{-3} & k \gg k_{\text{eq}} \end{cases} \tag{2.24}$$

2.3 Two-Point Correlation Function (2PCF)

A common way to study the distribution and clustering of galaxies is by computing the two-point correlation function (2PCF). If we have a distribution field $f(x)$, the 2PCF is:

$$\xi(\vec{x}_1,\vec{x}_2) \equiv \langle f(\vec{x}_1) f(\vec{x}_2)\rangle \tag{2.25}$$

If the field is homogeneous, the correlation function will only depend on the difference $\vec{r} \equiv \vec{x}_1 - \vec{x}_2$

$$\xi(\vec{r}) \equiv \langle f(\vec{x}) f(\vec{x}+\vec{r})\rangle \tag{2.26}$$

Furthermore, for a homogeneous and isotropic field this will only depend on $r \equiv |r|$.

The 2PCF of a random field can be understood as the average excess probability of finding two points in dV_1 and dV_2 separated by a distance $\vec{r}$, respect to the probability of finding those points at the same distance in a uniform random distribution with the same mean density ($\overline{n}$):

$$dP = \overline{n}^2(1+\xi(\vec{r}))dV_1dV_2 \tag{2.27}$$

If $\xi > 0$ the objects are correlated and we have clustering, if $\xi < 0$ the objects are anti-correlated and we have anti-clustering. If $\xi = 0$ the galaxies are randomly distributed, unclustered.

We can estimate the correlation function from the point process as:

$$1+\xi(r) = \frac{\text{\# pairs in the distribution at a distance r}}{\text{\# pairs expected for a random distribution}} \tag{2.28}$$

This estimator is perfectly fine as long as we can draw spheres of radius r around all the galaxies in our dataset. This is the case, for example, for data coming from an N-body simulation, in which the boundary conditions are periodic. However, in realistic cases the data lies inside a limited region of space, we just observe a patch of the sky and this introduces boundary effects that can affect the correlation function. The window function (also called mask) describes the area and depth of the observations. There are estimators that correct these effects, the optimal one is the Landy-Szalay estimator [79]:

$$\xi_{LS}(r) \equiv \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)} \tag{2.29}$$

where $DD(r)$ is the number of pairs in the data sample separated by a distance r, $RR(r)$ is the number of pairs in a random sample (with the same mask as the data) separated by a distance r, and $DR(r)$ is the cross correlation between both samples. All correlation functions computed in this thesis have been done using the code CUTE [80]. This code provides two implementations to compute two-point statistics: one for execution on shared-memory machines using OpenMP and one that runs on graphical processing units (GPUs) using CUDA.

The two-point correlation function makes the most efficient use of the data in a survey in terms of shot noise, since it uses all available galaxy pairs. The drawback of using the 2PCF is that the radial and angular scales, which contain complementary information, are mixed. It requires the use of a fiducial cosmology model to translate redshifts into distances. Angular $w(\theta)$ and radial $\xi_{\parallel}(\Delta z)$ correlation functions are based on observable quantities only $(\theta, \Delta z)$.

Angular Correlation Function

The angular two-point correlation function $w(\theta)$ can also be defined as the excess probability that two points in the differential solid angles $d\Omega_1$ and $d\Omega_2$ are separated by a distance θ, compared to a uniform distribution. The angular 2PCF is related to $\xi(r)$ by:

$$w(\theta) \equiv \langle \delta(\hat{n}_1)\delta(\hat{n}_2)\rangle = \int_0^\infty dz_1 \phi(z_1) \int_0^\infty dz_2 \phi(z_2) \xi(r_{12}, \mu) \tag{2.30}$$

where $\phi(z)$ is the redshift selection function given a probability of finding a redshift z_p given z, $\mathrm{P}(z_p|z)$:

$$\phi(z) \propto \frac{dN}{dz} \int_{z_{p1}}^{z_{p2}} dz_p \frac{dN_p}{dz_p} \mathrm{P}(z_p|z) \tag{2.31}$$

$r_{12} \equiv |\vec{x}_1 - \vec{x}_2|$ is the relative distance between two points, and μ the angle of the separation vector with respect to the line of sight:

$$\begin{aligned} r_{12} &\equiv \sqrt{x_1^2 + x_2^2 - 2x_1 x_2 \cos\theta} \\ \mu &\equiv \frac{(\vec{x}_1 - \vec{x}_2)\cdot(\vec{x}_1 + \vec{x}_2)}{|\vec{x}_1 - \vec{x}_2|\,|\vec{x}_1 + \vec{x}_2|} = \frac{|x_1^2 - x_2^2|}{\sqrt{x_1^4 + x_2^4 + 2x_1^2 x_2^2(1 - 2\cos^2\theta)}} \end{aligned} \tag{2.32}$$

2.4 Redshift Space Distortions (RSD)

In galaxy surveys we cannot separate peculiar velocities of galaxies from the Hubble flow. Then we see some features in the observations called Redshift Space Distortions (sec:1.2.7). If we assume that the patch of the Universe where clustering is measured is sufficiently far away, the line-of-sight is approximately constant (Kaiser approximation), the Kaiser effect can be modelled as [47]:

$$P_s(k) = \left(1 + \beta \mu_k^2\right)^2 P(k), \tag{2.33}$$

where the subscript s denotes redshift space, $\mu_k = \cos\theta$ in this space, and β is proportional to the growth rate, f, and to the galaxy bias $b(z)$:

$$\beta = \frac{f(z)}{b(z)} \equiv \frac{1}{b(z)} \frac{d\ln D_+}{d\ln a} \tag{2.34}$$

A useful expansion is [81]:

$$\left(1 + \beta\mu_k^2\right)^2 = \left[1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right] \mathscr{P}_0(\mu_k) + \left[\frac{4}{3}\beta + \frac{4}{7}\beta^2\right] \mathscr{P}_2(\mu_k) + \frac{8}{35}\beta^2 \mathscr{P}_4(\mu_k) \tag{2.35}$$

where $\mathscr{P}_l$ are the Legendre polynomials of degree l. This allow us to write $\xi(\vec{r})$ in terms of (r, μ) and expand the correlation function in terms of the Legendre polynomials as follows:

$$\xi(r, \mu) = \sum_l \xi_l(r) \mathscr{P}_l(\mu) \tag{2.36}$$

where ξ_l is defined as:

$$\xi_l(r) = \frac{i^l}{2\pi^2} \int dk P(k) k^2 j_l(kr), \tag{2.37}$$

and $j_l(kr)$ is the spherical Bessel function of order l. Thus, the redshift space correlation function including Redshift Space Distortions can be written as:

$$\xi_s(r, \mu) = b(z)^2 \left\{ \left[1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right] \mathscr{P}_0(\mu)\xi_0(r) + \left[\frac{4}{3}\beta + \frac{4}{7}\beta^2\right] \mathscr{P}_2(\mu)\xi_2(r) + \frac{8}{35} \mathscr{P}_4(\mu)\xi_4(r) \right\} \tag{2.38}$$

2.5 Higher Order Correlation Functions

The non-linear nature of gravitational dynamics leads, through mode-coupling effects, to the emergence of non-Gaussianity. So two-point statistics are not enough to characterize large-scale structure. They do not contain all the information available to constrain cosmological models. Higher-order statistics are needed to constrain galaxy-bias, primordial non-Gaussianities or break the degeneracies present in the measurements of two-point statistics. When initial conditions are not Gaussian, higher-order correlation functions are non-zero from the beginning and their evolution beyond LPT is non-trivial [82].

The absence of solutions of the equations of motion in the non-linear regime has motivated the search for consistent relations between correlation functions inspired by observations of galaxy clustering and the symmetries of dynamics. The most common example is the so-called hierarchical model for the connected J-point correlation function. Since non-linearities in the equations of motion are quadratic, gravitational instability generates connected higher-order correlation functions that scale as: $\xi_J \propto \xi_2^J$ at large scales, where $\xi_2 \ll 1$ and LPT applies. There is observational evidence that large-scale galaxy J-point correlation functions exhibit a hierarchical structure in both angular and redshift catalogues [83]. In the hierarchical model, all high-order correlations ξ_J can be expressed in terms of the two-point correlation function ξ_2 [83]:

$$\xi_J(r_1,...,r_J) = \sum_\alpha Q_{J,\alpha} \sum_{ab} \prod^{J-1} \xi_2(r_{ab}) \tag{2.39}$$

where Q_J are the hierarchical amplitudes. The same can be done in the angular projection:

$$w_J(\theta_1,...,\theta_J) = \sum_\alpha q_{J,\alpha} \sum_{ab} \prod^{J-1} w_2(\theta_{ab}) \tag{2.40}$$

The hierarchical 3-point correlation function is defined as:

$$\xi_3 \equiv \frac{\langle \delta(r_1)\delta(r_2)\delta(r_3)\rangle}{\xi(r_{12})\xi(r_{23}) + \xi(r_{12})\xi(r_{13}) + \xi(r_{13})\xi(r_{23})} \tag{2.41}$$

In contrast to the two-point correlation function, the three-point correlation function is not isotropic as it is sensitive to the shape of the large-scale structure. It therefore provides access to additional information.

2.6 Non-linearities

Galaxy surveys are increasingly improving their precision levels, reaching very low scales. At these low scales we are entering the extremely non-linear regime. Non-linearities in the dynamics correspond to couplings between modes of different wavelengths. The non-linear terms that are neglected in linear perturbation theory are not negligible any more. The Renormalized Perturbation Theory (RPT) [84] corrects the linear power spectrum for mode-coupling effects:

$$P(k,z) = G^2(k,z)P_{linear}(k) + P_{MC}(k,z) \tag{2.42}$$

where $G(k,z)$ is known as the propagator, and P_{MC} is the mode coupling term. The correlation function would be:

$$\xi(r,z) = \left[G^2 \otimes \xi_0\right](r,z) + \xi_{MC}(r,z) \tag{2.43}$$

In practice, a common approach to deal with non-linearities are numerical codes like `HALOFIT` [85]. This is a fitting formula based on a suite of high-resolution N-body simulations and decomposes the dimensionless power-spectrum $\Delta^2(k) = k^3 P(k)/(2\pi^2)$ as:

$$\Delta^2(k) = \Delta^2_Q(k) + \Delta^2_H(k) \tag{2.44}$$

The first term is called the two-halo term, that dominates at large scales, whereas the second term is referred to as the one-halo term, that is important at small scales.

2.7 Counts-in-Cells (CiC)

Counts-in-Cells (CiC) [49] is a method based on dividing a galaxy survey in cells of equal volume ($V_{\rm pix}$) and counting the number of galaxies in each cell ($N_{\rm gal}$). However, it is particularly useful to work with the density contrast, δ, in each pixel (eq:2.5), where ρ is the density and $\langle\rho\rangle$ the mean density:

$$\begin{aligned} \rho &= \frac{N_{\rm gal}}{V_{\rm pix}} \\ \langle\rho\rangle &= \frac{N^{\rm tot}_{\rm gal}}{V_{\rm tot}} \end{aligned} \tag{2.45}$$

The clustering can be studied from the moments of the density contrast distribution. It is a simple but powerful way to extract very valuable information from the distribution of galaxies and matter in the Universe.

2.7.1 CiC Moments

Most of the information is encoded in the moments of this distribution:

$$m_{j>1} = \langle\delta^j\rangle = \frac{1}{N_{\rm pix}} \sum_i (\delta_i - \langle\delta\rangle)^j \tag{2.46}$$

where $N_{\rm pix}$ is the number of pixels inside the survey. In practice, what we measure are the central moments of the angular counts:

$$m_J(\theta) \equiv \sum_{i=0}^{\infty} (i - \overline{N})^J P_i(\theta) \tag{2.47}$$

where $P_i(\theta)$ is the probability of finding i galaxies in a randomly selected cell of solid angle $A = 2\pi(1-\cos\theta)$ and $\overline{N} \equiv \sum_i i P_i$ is the average number of galaxies per cell. The galaxy density fluctuation in the cell is:

$$\delta_g = \frac{(i - \overline{N})}{\overline{N}} \tag{2.48}$$

and therefore:

$$m_J = \overline{N}^J \langle \delta_g^J \rangle \tag{2.49}$$

But since what we really want to obtain are the connected moments:

$$\mu_J \equiv \overline{N}^J \langle \delta_g^J \rangle_c \tag{2.50}$$

We have to correct what we measure, which is $\langle \delta_g^J \rangle$, subtracting the lower order contributions.

Connected moments

We can define a J-point correlation function of a density distribution ρ from its density fluctuations. The J-point correlation functions are [49]:

$$\xi_j(r_1, ..., r_j) \equiv \langle \delta_1, ..., \delta_j \rangle_c \tag{2.51}$$

where $\langle ... \rangle_c$ is the connected part of the expectation value. The connected part is the contribution to $\langle \delta_1, ..., \delta_j \rangle$ which does not include any conditional probability of lower order. Up to $J = 3$ we have $\langle ... \rangle_c = \langle ... \rangle$, but from $J = 4$ on they differ. For $J = 4$:

$$\langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle = \langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle_c + \sum_{ijkl} \langle \delta_i \delta_j \rangle \langle \delta_k \delta_l \rangle \tag{2.52}$$

as the conditional probability to have a pair i,j given a pair k,l contributes directly to the probability $\langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle$.

To estimate the connected graphs we introduce a moment-generating function:

$$M(t) = \sum_{j=0}^{\infty} \frac{m_j}{j!} = \langle e^{i\delta} \rangle \tag{2.53}$$

where $m_j = \left[\frac{d^j}{dt^j} M(t) \right]_{t=0}$ and the connected moments are:

$$\mu_j = \left[\frac{d^j}{dt^j} \log M(t) \right]_{t=0} \tag{2.54}$$

Up to $J = 6$ [86]:

$$\begin{aligned} \mu_2 &= m_2 \\ \mu_3 &= m_3 \\ \mu_4 &= m_4 - 3m_2^2 \\ \mu_5 &= m_5 - 10 m_2 m_3 \\ \mu_6 &= m_6 - 15 m_4 m_2 - 10 m_3^2 + 30 m_2^3 \end{aligned} \tag{2.55}$$

Hierarchical Clustering

The moments of the smoothed density field are the average of the J-point correlation functions in a cell of volume V_R [49]:

$$\overline{\xi}_J(R) = \langle \delta_R^J \rangle = \frac{1}{V_R^J} \int_{V_R} \xi_J(x_1, x_2) dx_1^3 dx_2^3 \tag{2.56}$$

If we do the average of the J-point correlation functions over a sphere of radius R, this means the J-order connected moments obey:

$$\overline{\xi}_J(R) = S_J [\overline{\xi}_2(R)]^{J-1} \tag{2.57}$$

where S_J are the rescaled moments:

$$S_J \equiv \frac{\langle \delta^J \rangle_c}{\langle \delta^2 \rangle^{J-1}} \tag{2.58}$$

Shot noise correction

The measurement of the statistical properties of the density field, in general, have a discrete character. Due to the discreteness, $\langle \delta_g^J \rangle_c$ is not a good estimator unless $\overline{N} \gg 1$, we have then to subtract the shot noise, $\delta_{\rm SN}$, which affects from $J = 2$ on. If there were not shot noise, the connected moments would equal $\overline{\xi}_J(R)$:

$$\overline{\xi}_J(R) = \langle \delta^J(R) \rangle_c - \delta_{\rm SN} \tag{2.59}$$

It is then natural to assume that the observed discrete distributions result from a Poisson realization of an underlying continuous field. This means that the probability of finding N points in a volume V at location $\vec{r}$ is given by $P_N^{Poisson} = [\overline{n}_g V(1 + \delta(\vec{r}))]$, where $P_N^{Poisson}$ is the probability of finding N objects in a Poisson process with expectation number $\overline{N} = \overline{n}_g V$, and $\overline{n}_g$ is the average number density.

$$P_N^{Poisson}(\overline{N}) \equiv \frac{\overline{N}!^N}{N!} e^{\overline{N}} \tag{2.60}$$

With this assumption, the moments of the discrete realizations can be related to the continuous ones. The generating function of the discrete field is related to that of the continuous field by $M_{Poisson}(t) = M(e^t - 1)$. Using the Poisson generating function, better estimators are obtained [86]:

$$\begin{aligned} k_2 &= \mu_2 - \overline{N} \\ k_3 &= \mu_3 - 3k_2 - \overline{N} \\ k_4 &= \mu_4 - 7k_2 - \overline{N} \\ k_5 &= \mu_5 - 15k_2 - 25k_3 - 10K_4 - \overline{N} \\ k_6 &= \mu_6 - 31k_2 - 90k_3 - 65k_4 - 15k_5 - \overline{N} \end{aligned} \tag{2.61}$$

where terms to the right of μ_J are the shot-noise correction.

Then, in our case what we have to do is just subtract the shot noise from what we measure $\langle \delta_g^J \rangle = \frac{m_J}{\overline{N}^J}$:

$$\begin{aligned}
\langle \delta^2 \rangle_c &= \frac{k_2}{\overline{N}^2} = \langle \delta_g^2 \rangle - \frac{1}{\overline{N}} \\
\langle \delta^3 \rangle_c &= \frac{k_3}{\overline{N}^3} = \langle \delta_g^3 \rangle - \frac{3}{\overline{N}} \langle \delta^2 \rangle_c - \frac{1}{\overline{N}^2} \\
\langle \delta^4 \rangle_c &= \frac{k_4}{\overline{N}^4} = \langle \delta_g^4 \rangle - 3\langle \delta_g^2 \rangle^2 - \frac{7}{\overline{N}^2} \langle \delta^2 \rangle_c - \frac{6}{\overline{N}} \langle \delta^3 \rangle_c - \frac{1}{\overline{N}^3} \\
\langle \delta^5 \rangle_c &= \frac{k_5}{\overline{N}^5} = \langle \delta_g^5 \rangle - 10\langle \delta_g^2 \rangle \langle \delta_g^3 \rangle - \frac{15}{\overline{N}^3} \langle \delta^2 \rangle_c - \frac{25}{\overline{N}^2} \langle \delta^3 \rangle_c - \frac{10}{\overline{N}} \langle \delta^4 \rangle_c - \frac{1}{\overline{N}^4} \\
\langle \delta^6 \rangle_c &= \langle \delta_g^6 \rangle - 15\langle \delta_g^2 \rangle \langle \delta_g^4 \rangle - 10\langle \delta_g^3 \rangle^2 + 30\langle \delta_g^2 \rangle^3 - \frac{31}{\overline{N}^4} \langle \delta^2 \rangle_c - \frac{90}{\overline{N}^3} \langle \delta^3 \rangle_c - \frac{65}{\overline{N}^2} \langle \delta^4 \rangle_c - \frac{15}{\overline{N}} \langle \delta^5 \rangle_c - \frac{1}{\overline{N}^5}
\end{aligned} \tag{2.62}$$

where $\overline{N}$ is in our case $\overline{N} = \frac{N_{\rm gal}^{\rm tot} A_{\rm pix}}{A_{\rm tot}}$, being $N_{\rm gal}^{\rm tot}$ the total number of galaxies, $A_{\rm tot}$ the total area, and $A_{\rm pix}$ the area of the pixel. Then the final moments would be:

$$S_J(\theta) \equiv \frac{\overline{w}_J(\theta)}{[\overline{w}_2(\theta)]^{J-1}} = \frac{\langle \delta^J \rangle_c}{\langle \delta^2 \rangle^{J-1}} \tag{2.63}$$

2.7.2 First order moments

The first moment of the distribution is the mean $m_1 = \langle \delta \rangle$. By construction, it is zero in the absence of numerical or boundary effects:

$$\langle \delta \rangle = \langle \frac{\rho}{\langle \rho \rangle} - 1 \rangle = 0 \tag{2.64}$$

The second moment of the distribution is the variance, $m_2 = \sigma^2$. Since the mean should be zero, then:

$$\sigma^2 = \langle \delta^2 \rangle - \langle \delta \rangle^2 = \langle \delta^2 \rangle \tag{2.65}$$

Using spherical cells we can obtain the third moment, the skewness, through the expansion:

$$\langle \delta^3 \rangle = \langle (\delta^{(1)} + \delta^{(2)} + ...)^3 \rangle = \langle (\delta^{(1)})^3 \rangle + 3\langle (\delta^{(1)})^2 \delta^{(2)} \rangle + ... \tag{2.66}$$

where the neglected terms are of higher order in LPT and the first term is zero for initial Gaussian conditions and then

$$\langle \delta^3 \rangle = \langle (\delta^{(1)})^2 \delta^{(2)} \rangle = 6a^4 \int d^3\vec{k}_1 \int d^3\vec{k}_2 P(k_1) P(k_2) F_2(\vec{k}_1, \vec{k}_2) \tag{2.67}$$

where $F_2(\vec{k}_1, \vec{k}_2)$ is a theory kernel coming from second-order perturbation theory [87]:

$$F_2(\vec{k}_1, \vec{k}_2) = \frac{5}{7} + \frac{1}{2} \frac{\vec{k}_1 \cdot \vec{k}_2}{k_1 k_2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{2}{7} \frac{(\vec{k}_1 \cdot \vec{k}_2)^2}{k_1^2 k_2^2} \tag{2.68}$$

Integrating over the angle between $\vec{k}_1$ and $\vec{k}_2$, The authors in reference [49] found that:

$$S_3 = \frac{34}{7} + \mathcal{O}(\sigma^2) \tag{2.69}$$

The skewness measures the asymmetry of the distribution, it can be positive, negative, or undefined. It is the most general estimator of non-Gaussianity. The skewness is:

$$S_3 = \frac{\langle \delta^3 \rangle}{\langle \delta^2 \rangle^2} \tag{2.70}$$

The same can be done for the fourth order moment, the kurtosis. Through the expansion $\langle \delta^4 \rangle = \langle (\delta^{(1)} + \delta^{(2)} + ...)^4 \rangle$, keeping the leading terms in perturbation theory and integrating in Fourier space it can be shown that $S_4 \approx \frac{60712}{1323}$ [49]. The kurtosis assesses whether the distribution is more peaked or flatter than a Gaussian distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. The kurtosis is:

$$S_4 = \frac{\langle \delta^4 \rangle_c}{\langle \delta^2 \rangle^3} \tag{2.71}$$

Filtered moments

These computations only apply in local, unfiltered density fields. In practice, fields are always observed at a finite spatial resolution, both in observations and in simulations. In terms of the smoothed fluctuations, the volume-averaged J-point correlation functions are:

$$\overline{\xi}_j \equiv \langle \delta_W^j \rangle_c = \frac{1}{V_W^j} \int_s dr_1 ... dr_j W(r1) ... W(r_j) \xi_j(r_1 ... r_j) \tag{2.72}$$

In the case of $J = 2$:

$$\overline{\xi}_2(V_W) = \frac{1}{2\pi^2} \int_0^\infty dk\, k^2 P(k) W^2(k) \tag{2.73}$$

Filtering means convolving a window function with the density field. Equation (2.67) would become

$$\langle \delta_R^3 \rangle = 3 \langle (\delta_R^{(1)})^2 \delta_R^{(2)} \rangle = 6a^4 \int d^3\vec{k}_1 \int d^3\vec{k}_2 P(k_1) P(k_2) W_3(k_1 R) W_3(k_2 R) \times F_2(\vec{k}_1, \vec{k}_2) W_3((\vec{k}_1 + \vec{k}_2) R) \tag{2.74}$$

where $W_3(k)$ is the 3D filtering function in Fourier space. For a top-hat window function

$$W_3(k) = \frac{3}{k^3} (\sin(k) - k\cos(k)) \tag{2.75}$$

what is obtained is [88]:

$$S_3 = \frac{34}{7} + \frac{d \log \sigma^2(R)}{d \log R} \tag{2.76}$$

The skewness thus depends on the power spectrum shape, mainly at the filtering scale. For a power-law power spectrum, $P(k) \propto k^n$, this scale dependence disappears in the linear regime and it follows that:

$$S_3 = \frac{34}{7} - (n+3) \tag{2.77}$$

This can be used to obtain the hierarchical amplitude Q_3 [83]:

$$S_3 = 3\eta(n) Q_3 = \frac{34}{7} - (n+3)$$
$$\eta(n) = \frac{(3-n)^2}{12(1+n)^2} \left[-\frac{2(-1+n^2+2n^3)}{3-8n-4n^2} + \frac{4^n (1+n) \sqrt{\pi} \Gamma(2-n)}{\Gamma(\frac{3}{2}-n)} - \frac{4^n \sqrt{\pi} \Gamma(2-n)}{\Gamma(\frac{5}{2}-n)} \right] \tag{2.78}$$

Higher hierarchical amplitudes Q_J can be obtained from the hierarchical relation [83]:

$$Q_J = \left(\frac{4Q_3}{J}\right)^{J-2} \frac{J}{2J-2} \tag{2.79}$$

For the fourth order moment a similar computation can be done. For a power-law power spectrum the authors from reference [88] found that:

$$S_4 = \frac{60712}{1323} - \frac{64}{3}(n+3) + \frac{7}{3}(n+3)^2 \tag{2.80}$$

As skewness is induced by gravity, it depends on cosmological parameters. However, this dependence is so small that nowadays it can be ignored within errors [89]:

$$S_3 = \frac{34}{7} + \frac{6}{7}(\Omega_m^{0.03} - 1) - (n+3) \tag{2.81}$$

For other non-standard model equations like quintessence the same is found in reference [90].

To avoid Redshift Space Distortions in spherical cells, [83] change the shape of the volume to cells with the shape of conic sectors of a sphere between radii $d < r < D$. They found that for a power-law $\xi_2(r) = (r_0/r)^\gamma$:

$$S_N^C = N^{N-2} Q_N \frac{I_0^{N-2} I_{N-1}}{I_1^{N-1}} \tag{2.82}$$

where

$$\begin{aligned} I_k &= \frac{1-(d/D)^{3(k+1)-k\gamma}}{3(k+1)-k\gamma} \int_0^1 z dz F^k(z) \\ F(z) &= \int_0^1 x dx \int_0^{2\pi} d\phi (z^2 + x^2 - 2xz\cos(\phi))^{(1-\gamma)/2} \end{aligned} \tag{2.83}$$

2.7.3 Angular CiC

Previous 3D CiC definitions can be easily generalized for angular aperture cells. Angular CiC is based on dividing a galaxy survey in cells of equal area ($A_{\rm pix}$) and counting the number of galaxies in each cell ($N_{\rm gal}$). The density contrast in each pixel is given by equation (eq: 2.5) where the density is now $\rho = \frac{N_{\rm gal}}{A_{\rm pix}}$, and $\langle\rho\rangle = \frac{N_{\rm gal}^{\rm tot}}{A_{\rm tot}}$.

In the angular case, where the data analysis simplifies and we have a better signal-to-noise ratio properties, we have that:

$$\overline{w}_j(\theta) \equiv \frac{1}{A^j} \int_A dA_1 ... dA_j w_j(\theta_1 ... \theta_j) = \langle \delta^j(\theta) \rangle_c \tag{2.84}$$

where $A = 2\pi(1 - cos\theta)$ is the solid angle of the cone, $dA_J = sin\theta_J d\theta_J d\phi_J$ and $\delta(\theta)$ are the fluctuations inside the cone.

We can define the analogous relationship of equation (2.57), for the area-averaged angular correlations $\overline{w}_J(\theta)$:

$$S_J = \frac{\overline{w}_J(\theta)}{[\overline{w}_2(\theta)]^{J-1}} \tag{2.85}$$

In this case the moments would be the average of the J-point angular correlation functions in a cell of area A:

$$\langle \delta^J \rangle = \frac{1}{A^J} \int_A w_J(\theta_1, \theta_2) d\Omega_1 d\Omega_2 \tag{2.86}$$

For $J = 2$:

$$\overline{w}_2(\theta) = \frac{1}{A^2_{\text{pix}}} \int w_2(\theta) d\Omega_1 d\Omega_2 \tag{2.87}$$

2.7.4 Previous CiC measurements

CiC has been widely studied in the different galaxy surveys (APM, CFHTLS, SDSS, 2dFGRS, 2MASS, VIMOS,... [86, 91, 92, 93, 94, 95]). The first systematic study of galaxy clustering was carried out in the 1970s by Peebles and his collaborators [49]. They confirmed the power-law behaviour of the angular two-point correlation function and the hierarchical scaling with poorly calibrated photographic plates. After some years and some technological developments, Automatic Plate Measuring Machine (APM) was one of the first surveys to measure high order correlation functions and high order moments. The authors in reference [96] found that hierarchical amplitudes are roughly constant, up to $J = 9$, and decrease slowly for larger scales. A later analysis confirmed their results in agreement with SDSS within a 90% confidence level [96].

Despite the large amount of measurements, just a few of them have studied the redshift evolution of the hierarchical moments. The authors in reference [91] find with the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) that at non-linear scales, the amplitude of the hierarchical moments increases with redshift. However, a more accurate error analysis demonstrates that their results are also consistent with no redshift evolution within 2σ [91]. Using spectroscopic redshifts from the VIMOS-VLT Deep Survey (VVDS), The authors from reference [95] measured the evolution of $\overline{\xi}_2$ and S_3 over the redshift range $0.7 < z < 1$. They found that the redshift evolution in this interval is consistent with predictions of first and second-order LPT. With the Sloan Digital Sky Survey (SDSS) the authors from [97] found that cosmic variance in the SDSS causes CiC distributions in different quadrants to differ from each other by up to 20%. Many surveys have measured J-point galaxy area-averaged angular correlation functions up to $J = 7$ finding roughly constant behaviour consistent with the hierarchical model at scales $0.3\text{Mpc/h} < \text{r} < 40\text{Mpc/h}$ [92, 93, 94].

Another way of using CiC is keeping the number of galaxies constant and varying the volume of the cells. This is used to study the void probability function (VPF) [98], where the number of galaxies is kept at zero.

2.7.5 Lognormal

We cannot predict the specific locations of galaxies around us, as their distribution is statistical. What we can do is fitting a stochastic model to the cosmological density field. Stochastic models are used in physics to describe phenomena and discriminate between theory and observations, even if they are not physically motivated. However, these models are completely specified statistically and do not violate common-sense conditions. The authors in reference [99] study the lognormal (LN) random field to describe the continuous density field. The advantage of this model is that it is fully specified statistically, it always has $\rho > 0$, and becomes arbitrarily close to Gaussian statistics at early times. Inflationary models suggest that linear density perturbations are described by Gaussian statistics. A Gaussian random field, however, can only be a model for the linear density field, i.e., in the limit of zero fluctuation amplitude. The Gaussian random field is suitable for describing additive processes, thanks to the central limit theorem. The analogous for multiplicative processes would be the lognormal random field. If we can write a process Y as a product of independent terms, then $\ln(Y)$ is normally distributed, so Y is lognormally distributed. As Gaussian models can be used to describe linear processes, lognormal models can be used to describe non-linear ones.

The authors from reference [99] show how the continuous density field can be written as a multiplicative process, then it can be modelled by a lognormal (LN) random field. The continuity equation for the matter flow is:

$$\frac{\partial \rho}{\partial t} + 3\left(\frac{\dot{a}}{a}\right)\rho + \frac{1}{a}\nabla\cdot(\rho v) = 0 \tag{2.88}$$

If we change time coordinate to conformal time ($dt = ad\tau$) now the density variable is $\rho = \rho a^3$ and the continuity equation is:

$$\frac{1}{\rho}\frac{\partial \rho}{\partial \tau} = -(\nabla\cdot v) \tag{2.89}$$

If the initial peculiar velocity field is Gaussian then so is $(\nabla\cdot v)$. If it continues to grow linearly then the peculiar velocity divergence stays Gaussian and scales with τ: $(\nabla\cdot v)_\tau = (\nabla\cdot v)_{\tau_0}(\tau/\tau_0)$. Therefore

$$\rho(x) = \rho_0 \exp\left[\varepsilon(x)\tau^2\right] \tag{2.90}$$

where ε is a Gaussian random field $\varepsilon = -\frac{1}{2}(\nabla\cdot v)_{\tau_0}\tau_0^{-1}$. Then, assuming linear velocity fluctuations $X = \ln(Y)$, the density distribution can be modelled by a lognormal Y. It is a crude model as it does not take into account the geometry of the local velocity field. But it seems rather natural and it is one of the simplest ways of defining a fully self-consistent random field with $\rho > 0$.

2.8 Bias

Most of the matter in the Universe is made of dark matter, which gravitates but does not interact with light. What we really observe in galaxy surveys is the galaxy distribution,

not the matter distribution. Galaxies are a small part of the total amount of matter in the Universe, but they are tracers of dark matter. If we want to infer the mass distribution from galaxy surveys it is crucial to understand the connection between the galaxy and the dark matter distribution. Both, baryons and dark matter structures grow around primordial overdensities. Structure formation is hierarchical, in the sense that small structures form first and larger structures form later. Galaxies are expected to follow dark matter potential wells [100]. Then, we expect the large-scale structure of galaxies to be correlated with dark matter.

Galaxy biasing was seen for the first time analysing the clustering of different populations of galaxies [101, 102]. The theoretical relation between galaxy and mass distributions was suggested by Kaiser [103], and developed by authors in reference [104]. Since then, many different prescriptions have arisen [105, 106, 107, 108, 109, 110]. However, there is no generally accepted framework for galaxy biasing yet. While the galaxy and the dark matter distribution are related, the exact relation depends on galaxy formation [111], galaxy evolution [112, 113, 114] and selection effects. Galaxy bias depends on many aspects, such as scale, galaxy properties(luminosity, colour,...), sample selection, redshift,... Bias depends strongly on the environment. Using dark matter simulations, the authors in reference [115] show how halo bias is determined by local density and not by halo mass. In several studies the different behaviour of early-type galaxies and late-types at both small and large scales is found [92, 116, 117, 118]. To have a good estimate of the real matter distribution it is convenient to use a galaxy sample as homogeneous as possible. Understanding the relation between galaxies and matter is essential for the measurements of cosmological parameters [119]. The uncertainties in this relation strongly increase the errors in the dark energy equation of state or gravitational growth index from future galaxy surveys [23].

Bias also depends on redshift [120, 113]. The bias is a measure of how well galaxies trace the dark matter, where a higher bias means a worse tracer. At high redshifts, galaxies formed in regions where the density was highest, therefore tracing poorly the total matter distribution. Over time, they started forming in less dense areas as well thus mapping dark matter more accurately. This manifests as a decrease of galaxy bias to lower redshifts.

2.8.1 Linear Bias

With the linear bias $b(z)$ approximation, we can relate the matter fluctuations δ_m and the fluctuations in the galaxy distribution δ_g:

$$\delta_g = b(z)\delta_m \tag{2.91}$$

This relation is a good approximation on large scales. In the linear approximation, up to scalings, all statistical properties are preserved by the biasing, and the observed galaxy properties do reflect the matter distribution. However, in the general case, it is highly unlikely that the relation is both local and linear. Non-local dependencies might come from some properties as the local velocity field, or derivatives of the local gravitational potential [105, 121].

With this local bias model, one can compute the two and three-point biased correlation functions, to find ([105, 122]):

$$\xi(r_{12}) \equiv \langle \delta(r_1)\delta(r_2)\rangle \simeq b^2 \xi_m(r_{12}) \tag{2.92}$$

2.8.2 Non-Linear Bias

Linear bias might be reasonable as long as we consider just two-point statistics. However, when non-Gaussianities are taken into account, linear bias fails to be a good description. If we want to go to higher orders we can assume that the (smoothed) galaxy density can be written as a function of the mass density and expand it as a Taylor series (assuming a local relation) [122]:

$$\delta_g = f(\delta) = \sum_{k=0}^{\infty} \frac{b_k}{k!}\delta_m^k \tag{2.93}$$

The linear term $b_1 = b$ is the usual linear bias. To have $\langle \delta_g \rangle = 0$ we must fix b_0:

$$b_0 = -\sum_{k=2}^{\infty} b_k \langle \delta_m^k \rangle / k! \tag{2.94}$$

The value of b_0 is irrelevant for the higher moments. In general, to predict higher order correlations to J-order, the local relation has to be expanded to order J-1. The moments of the density contrast can be related to this non-linear approach. Fry [105] obtained for the higher order moments the following:

$$\begin{aligned} S_3 &= b^{-1}(S_{3\ m} + 3c_2) \\ S_4 &= b^{-2}(S_{4\ m} + 12c_2 S_{3\ m} + 4c_3 + 12c_2^2) \end{aligned} \tag{2.95}$$

where $c_k = b_k/b$ for $k \geq 2$.

Galaxy formation not only depends on gravity, but also on non gravitational physics. At sufficiently large scales structure formation is dominated by the gravitational interaction, and therefore, galaxies can probe the underlying matter density. The authors from [110] show that the local bias is consistent for scales larger than $R > 30 - 60\mathrm{Mpc/h}$. In the same scale, the authors in reference [110] found that the values of b_1 and b_2 in MICE simulation approach a constant value. It has also been found that higher-order galaxy bias is non-negligible [92]. Using DR5 from SDSS the authors from [123] found that early-type galaxies have a larger c_2 than late-type galaxies.

Moreover, the authors in reference [124] found that non-local terms are responsible for the overestimation of linear bias from the three-point correlation observed by [125, 126, 110]. Although they do not expect non-local bias to significantly affect second-order statistics.

Due to the increasing amount data from large-scale galaxy surveys, errors of statistical measurements are rapidly decreasing. This high level of precision requires the same level of accuracy in modelling the different observables. At second order the non-local terms can be expressed as [124]:

$$\delta(x) = b_1 \delta_m(x) + \frac{b_2}{2}(\delta_m^2(x) - \langle \delta_m^2 \rangle) + \gamma_2 \mathscr{G}_2(x) \tag{2.96}$$

where γ_2 is the non-local bias parameter [127, 128]. This non-local component, $\mathscr{G}_2$, comes from the divergence of the velocity field and mode couplings.

2.8.3 Previous bias measurements

The complexity of galaxy bias is reflected in the literature, where many different approaches have emerged. There are several studies of galaxy bias in simulations (see e.g. [129, 130, 110, 131, 132]) and with different methods [133, 87, 110, 134, 125, 126, 124]. Measuring the bias from observations requires usually strong assumptions on cosmological parameters or dark matter clustering [135, 136, 137, 138, 139, 140, 141, 142].

Although the different estimators appear to measure the same bias, the results can be effected by stochasticity of the bias and projection effects between δ_g and δ_m. Moments computed from CiC are closer to the local relation in that they are both smoothed quantities, so one would expect a better agreement for them. But they suffer from shot noise and stronger non-linear effects (as they include clustering on all scales smaller than the smoothing radius). The two and three-point correlation functions do not suffer from shot-noise and can better separate the effect of different scales. Moreover, the three-point correlation function provides different information than the skewness. Both are related third order statistics, but the three-point correlation function also gives shape information. Different estimators and different order moments take non-linear and non-local terms in different ways [124].

The bias can be obtained in two steps, first the bias between halos and dark matter, and then the bias between halos and galaxies. According to the halo model, the dark matter field is characterized by dark matter haloes, and these haloes, although they can have different size and mass, present a universal profile [143]. According to the halo model, galaxies are formed inside dark matter haloes. A common way to relate halos and galaxies is with the so-called Halo Occupation Distribution (HOD) model. These models populate halos with the probability that a halo of a given mass M hosts N galaxies, $P(N|M)$. The expected galaxy occupation can be parametrized in terms of central and satellite galaxies, usually with three or four parameters [144, 145, 146]. These parameters, then, can be fitted from observations in order to learn about galaxy formation and the occupation of galaxies in haloes.

Another common way to obtain the bias is from the mass function. There are different approaches to model the mass function, for example the Sheth Tormen [108], or the Press-Schechter [111]. The authors in references [107, 109] have studied their accuracy.

Many of the previous studies are done with dark matter halos, however it is not straightforward to apply these results directly to galaxies. They can be applied to galaxies in the limit where halo biasing resembles galaxy biasing or in the limit where observations are good tracers of the halo distribution (i.e., for galaxy groups or clusters).

3. The Dark Energy Survey (DES)

The Dark Energy Survey [1] (DES) [59] is a photometric survey, designed to probe the origin of the accelerating universe and help uncover the nature of dark energy. This collaboration has an extremely sensitive 570-Megapixel digital camera, DECam, mounted on the Blanco 4-meter telescope (Fig. 3.1) at Cerro Tololo Inter-American Observatory high in the Chilean Andes. Starting in September of 2012 and continuing for five years, DES will survey 5000 square degrees of the southern sky in 5 optical filters (*grizY*). DES has completed its fourth observing season in February 2017. DES is an international collaboration, with over 500 scientists from 28 institutions and consortium in the US, Chile, the UK, Spain, Brazil, Switzerland, and Germany.

In this chapter I describe the survey: the science (sec: 3.1), the observations (sec: 3.2), the camera (sec: 3.3), the survey strategy (sec: 3.4), the data management DESDM (sec: 3.5), other science besides dark energy that can be achieved with the data (sec: 3.6), photometric redshifts (sec: 3.7), currently available data (sec: 3.8) and simulations (sec:3.9).

3.1 Science

The Dark Energy Task Force joined in 2005 to study and establish a joint dark energy program [2]. DES established itself as a Stage III survey. What makes DES special is that it combines four probes of dark energy in a single experiment:

- Type Ia Supernovae (SN) (sec: 1.2.2)
- Baryon Acoustic Oscillations (BAO) (sec: 1.2.4)
- Galaxy clusters (GC) (sec: 1.2.6)
- Weak Gravitational Lensing (WL) (sec: 1.2.4)

The first two (SN and BAO) constrain the expansion of the universe as a whole and are referred to as 'purely geometric'. The latter two (WL and GC) measure both the expansion of the universe and the growth of large-scale structures. The main systematic error sources

[1] `www.darkenergysurvey.org`

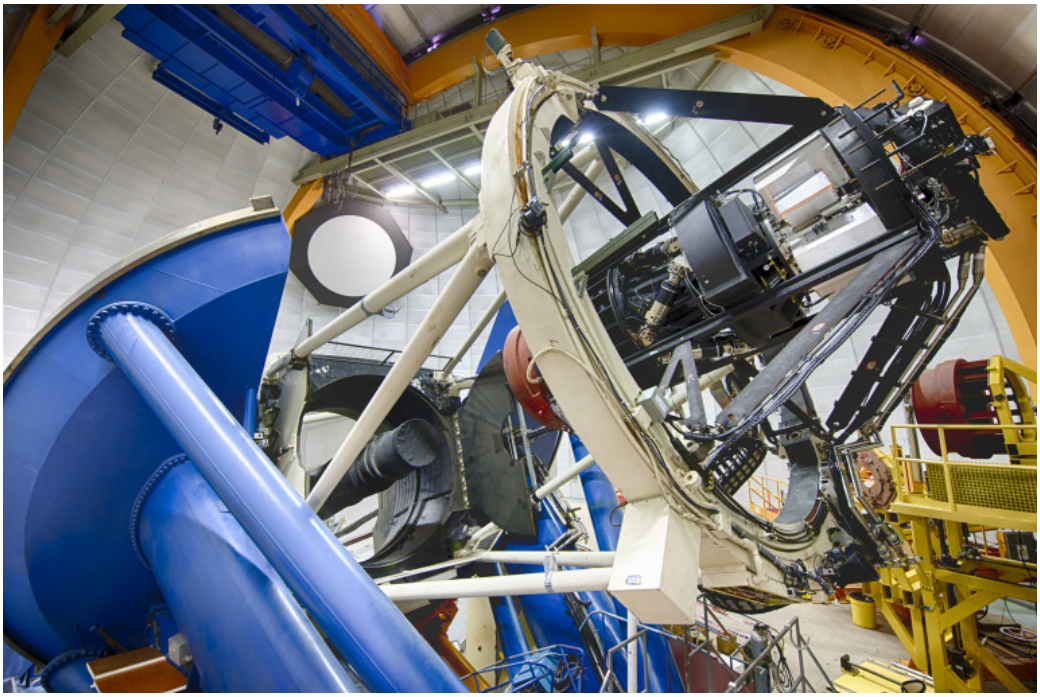


Figure 3.1: The Blanco 4-meter telescope at Cerro Tololo Inter-American Observatory high in the Chilean Andes.

for each method, ordered approximately from most to least important, along with the primary methods for controlling them are listed in Table 3.1.

These four probes can be combined to constrain the dark energy parameters, w_0 and w_a. The figure of merit expected to get at the end of the survey is in Figure 3.2. It is expected that the uncertainty on w_0 will be only a few percent for each probe. DES will improve the Dark Energy Task Force figure of merit by a factor of 3-5 over stage II projects [59].

The DES science is coordinated by a Science Committee comprised of thirteen Science Working Groups (SWGs). Core DESWGs include large-scale structure, clusters, weak lensing and supernovae Ia. Additional SWGs which focus on the primary science are photometric redshifts, spectroscopy, simulations, and theory & combined probes. The Non-Dark Enetgy SWGs focus on Milky Way science, galaxy evolution, strong lensing, quasars, and transients moving objects (sec: 3.6).

3.2 Observations

The collaboration formed in 2004 to design and build The Dark Energy Camera (DECam) (sec: 3.3) from 2004 to 2011. In exchange for the camera, the DES collaboration was allocated 105 nights per year of telescope time for 5 years. The first light was obtained in September 2012, followed by commissioning of DECam [149], and science verification (SV) observations from November 2012 to February 2013. Timing for observations can be seen in Table 3.2. Y4 was completed in February 2017, Y3 data is being analysed now, and Y1 and Y2 science papers will come out soon. The observations from Y1 to Y3 are

Method	Dominant Systematic Errors	Primary Controls
Clusters	Sample selection Mass-observable relation	SZE + optical cluster selection; simulations Self-calibration; statistical WL masses
Weak Lensing	Multiplicative shear Additive shear Photo-z biases Small-scale power spectrum	Measurement algorithm; shear vs. gal. size PCA; active focus; wave-front sensing; alignment control Spectroscopic calibration sets Null small-scale power; high-res. simulations
Angular Clustering	Bias prescription errors Large-scale photometric calibration errors Photo-z biases	Angular bispectrum;clustering by type Calibration strategy; clustering by color; angular sub-samples Spectroscopic calibration sets
Supernovae Ia	SN evolution Photometric errors Extinction Photo-z errors and biases	Low and high z SNe comparison Calibration strategy; artificial SNe SN color and host galaxy information SN spectroscopic calibration sub-samples

Table 3.1: Dominant sources of systematic error and methods for controlling them [59].

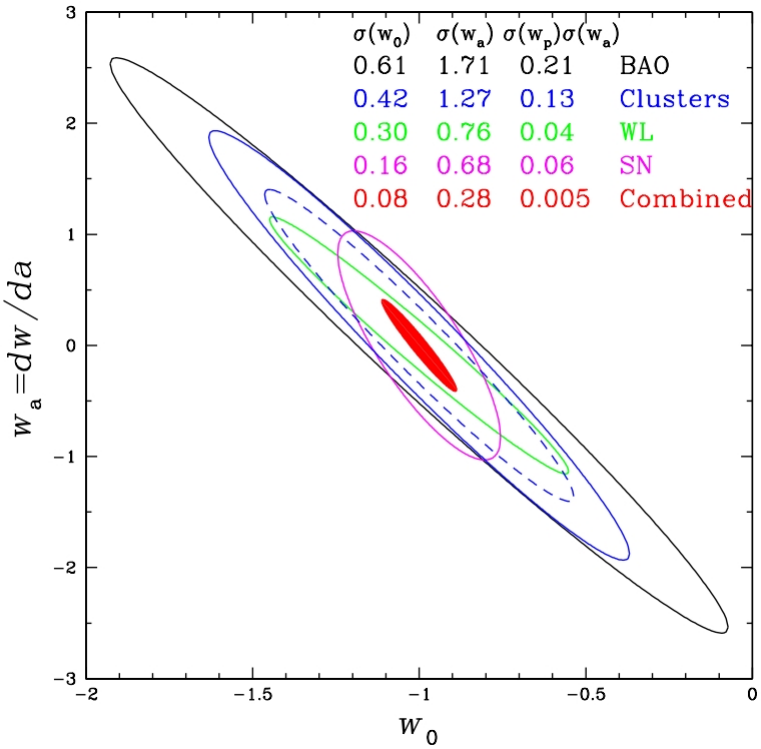


Figure 3.2: 68% joint likelihood contours for the different DES probes and combined. From the outside in the solid lines are from BAO, cluster counts with $\sigma_8 = 0.75$, cluster counts with $\sigma_8 = 0.9$ (dashed), weak lensing and Supernovae. The filled contour is obtained when the probes are combined. The inset table in the Figure shows the accuracy of the measurement on w_0 and w_a and the inverse figure of merit $\sigma(w_p) \times \sigma(w_a)$. Figure from [59].

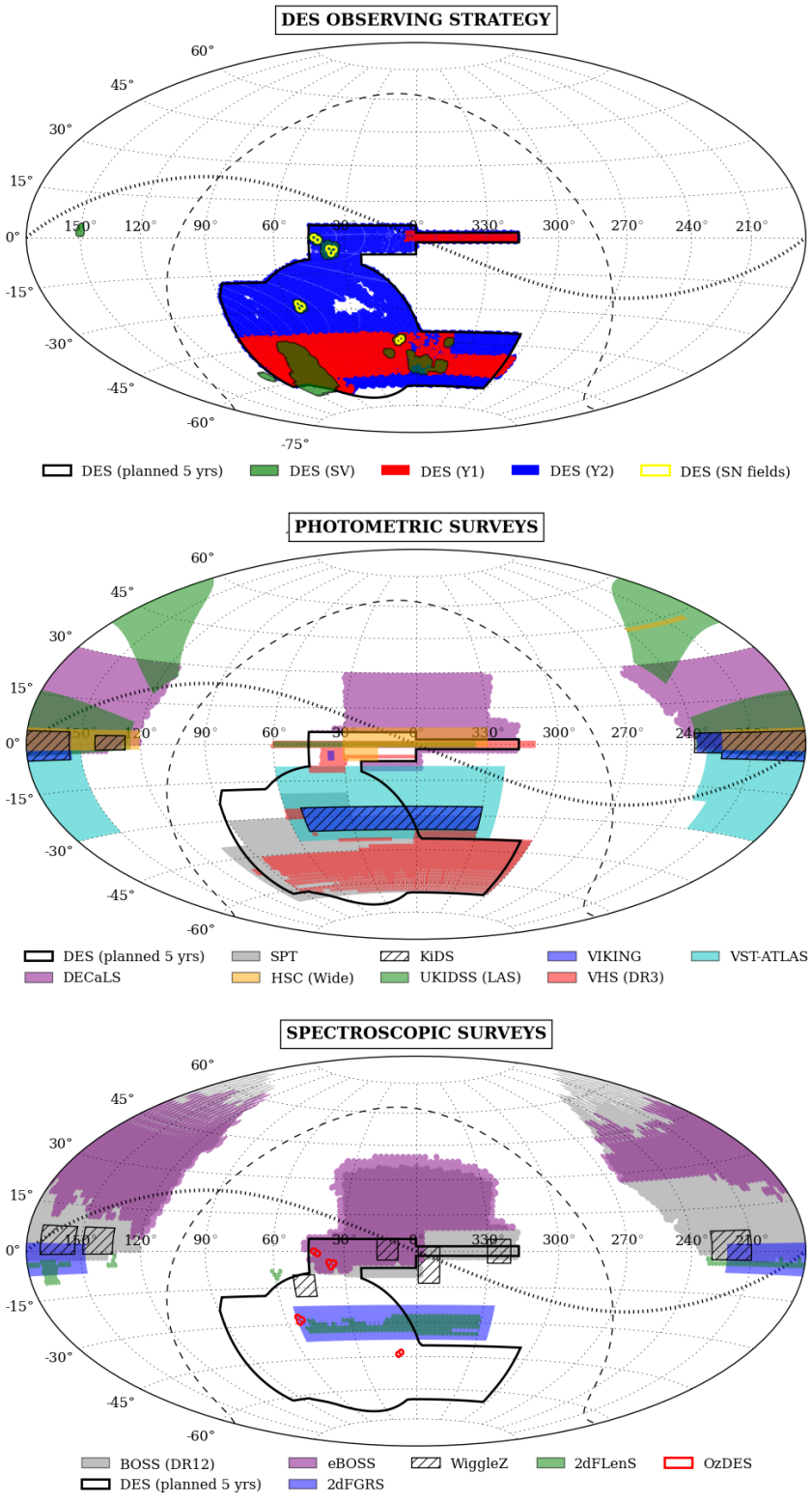


Figure 3.3: DES and some selected completed or ongoing surveys (as of December 2015). This is a Hammer projection in equatorial coordinates, with the dashed and dotted lines indicating the Galactic plane and the ecliptic plane, respectively [147]. Top: DES survey footprint for the SV, Y1, Y2 and the final 5-yrs survey; Middle: with other photometric surveys; Bottom: with other spectroscopic surveys.

Season	Start Date	End Date	2^{nd} Half Nights	Full Nights	1^{st} Half Nights
"Y1"	Aug. 31, 2013	Feb. 09, 2014	0	91	28
"Y2"	Aug. 15, 2014	Feb. 15, 2015	10	80	41
"Y3"	Aug. 04, 2015	Feb. 12, 2016	32	73	39
"Y4"	Aug. 2016	Feb. 2017			

Table 3.2: Scheduled start and end dates for DES observing, and the number of half-nights and full nights for Y1 to Y3 [148].

described in reference [148]. DES lost $\sim 30.3\%$ of their possible observing time due to bad weather, diagnosed as a strong El Niño Southern Oscillation. The DES footprint for the different seasons is shown in Figure 3.3, along with other photometric and spectroscopic surveys. The wide field survey will be 5000deg^2, and 30deg^2 area for detection of supernovae.

Twelve months after raw survey images are taken, reduced, calibrated and processed through the DES Data Management (DESDM) System (sec: 3.5), they are released and made public through the NOAO Science Archive (NSA) [2]. In addition, DES will make two public releases of co-added images and catalogues produced and served by DESDM, one based on the first two seasons of data, the second based on the full survey data set. DES Science Verification (SV) photometric sample (sec: 3.8) is already public [3].

DES has made good use of the initial data samples, with discoveries of new solar system objects [150], discoveries of dwarf galaxies and structure in the Milky Way [151, 152, 153], systematic studies of galaxy clustering [142], galaxy clusters [154], weak gravitational lensing [155, 156], and our first cosmological results [157]. There are currently more than 120 papers in some state of publication. By its completion the survey is expected to have generated a catalogue of 300 million galaxies with photometric redshifts, 100 million stars, 100000 galaxy clusters and about 3000 type Ia SNe.

The DES observations are staffed at the telescope by collaboration members. I had the chance to go to Cerro Tololo for DES observing the first week of January 2016 for Y3. There are three observing roles during full nights. 'Observer 1' controls the camera through the data-acquisition interfaces and executes the nightly program. This observer ensures that images are being recorded, pays attention to the alarms, and solves routine problems. 'Observer 2' performs quality control procedures, checks the exposures and ensures that the image quality is as expected given the current conditions. The 'Run Manager' is the lead observer and is responsible for ensuring that the two other observers perform their roles. I went observing a half-night observation day, when usually only two observers are needed with the Run Manager taking on one of the roles.

By now, the camera and the telescope have been widely used and tested by DES and the science community. The DESDM has automated codes and systems for everything,

[2]Public data release `http://des.ncsa.illinois.edu/releases`

[3]This sample is available at `https://opensource.ncsa.illinois.edu/confluence/display/DESDM/DES+SVA1+Data+Products`

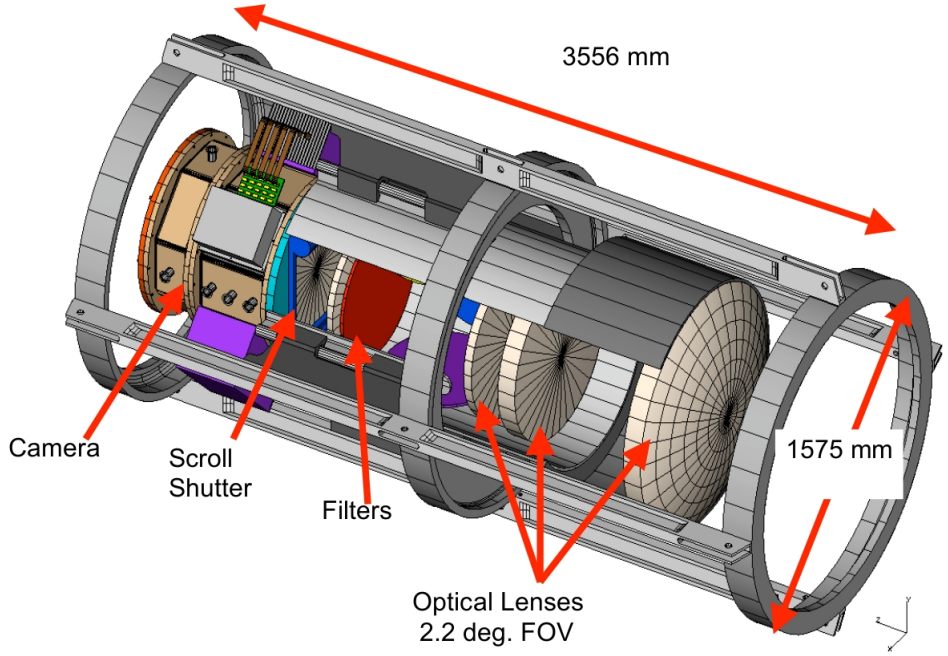


Figure 3.4: DECam Reference Design [59]

you just have to make sure everything works fine and you do not hear the chicken alarm. In the case of warning or alarm, procedures have been established. Besides, support is available to the observers through the CTIO Telescope Operator (on hand), the CTIO Observer Support Specialist (on-site), the CTIO Instrument Scientist (typically by phone), and the DES Operations Scientist & Support Team (by internet connection).

The daily operations cycle starts with a meeting in the afternoon to discuss any technical or procedural problems that occurred during the previous night, and discuss what is going to be observed during the night. The period before twilight is used for calibrations and to establish the basic functionality of the instrument. An hour before sunset the telescope operator will open the dome. At minus 10 degree twilight (roughly 40 minutes after sunset) the observers execute three standard star field exposure scripts to calibrate DES filter bands, and characterize that night's instrumental and atmospheric response. At minus 12 degree twilight (roughly 48 minutes after sunset) the observers begin the 'Observing Tactician' (OBSTAC) observations. OBSTAC selects the priority fields to observe every night. These continue during the night until minus 10 degree morning twilight, for standard star observations and calibrations. At the end of each night the observers create night summaries listing weather conditions, problems encountered, the fraction of time lost to each, the expected plan, the conditions and the accomplishments.

3.3 DECam

The Dark Energy Camera (DECam) [149] was built from 2004 to 2011, during this time many upgrades and enhancements were applied to the telescope and dome in order to maximize DECam's potential. DECam is a 570 Mpixel, 2.2-degree field-of-view camera.

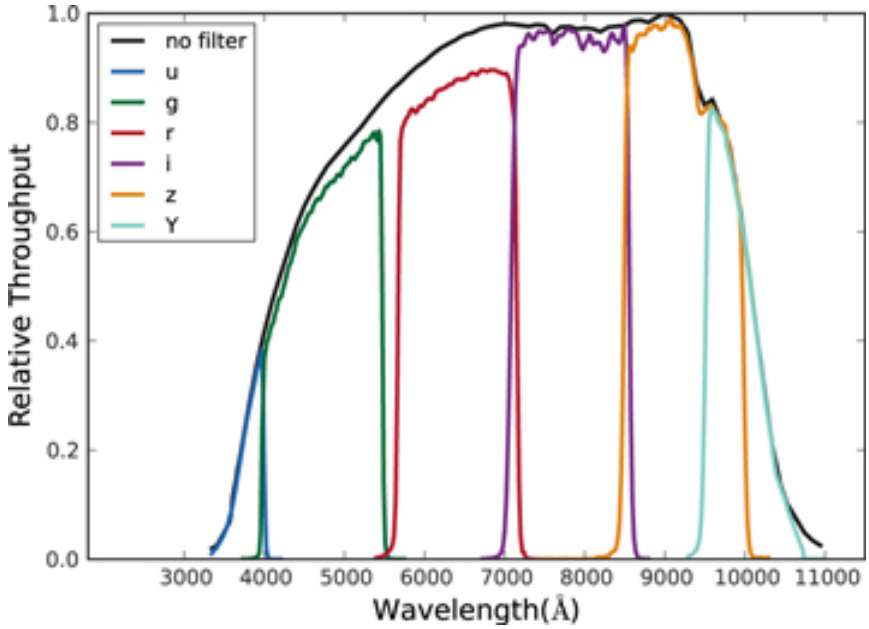


Figure 3.5: Throughput as a function of wavelengths of the DECam optical train, including the various filters DES uses. These throughputs are calculated relative to the use of no filter at 9000 Angstroms [149].

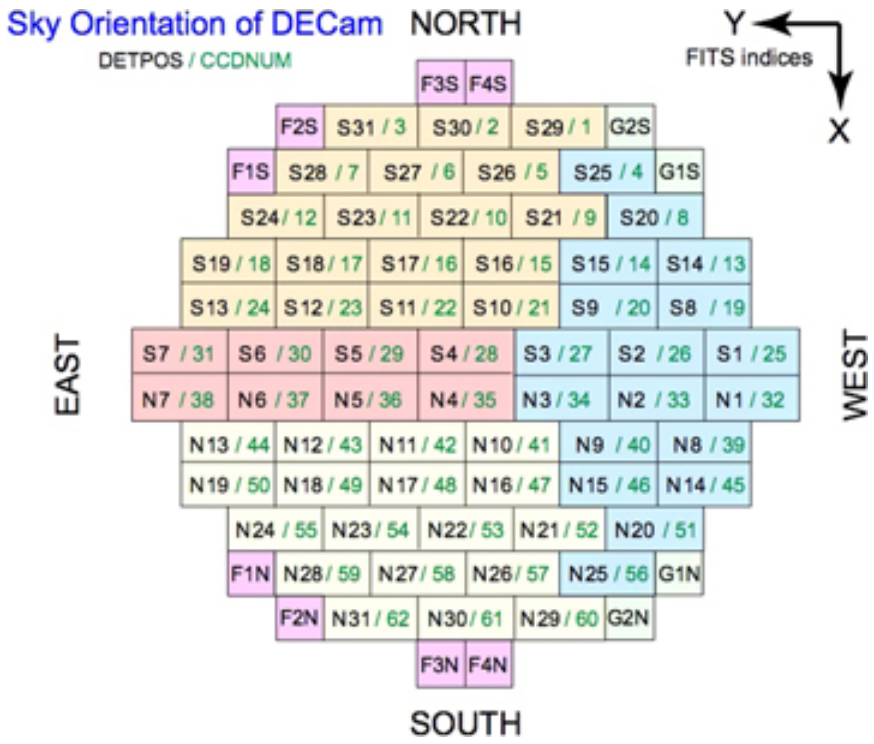


Figure 3.6: The DECam focal plane consists of 62 2048×4096 science CCDs distributed over a 2.2 degree diameter field of view. The DES imager has 3deg^2 of active area corresponding to 520 million pixels, each subtending ∼0.27 arcsec. Guider and auto-focus CCDs are marked with G and F [149].

It is currently installed and operating as a survey and community instrument at the Prime Focus of the Victor M. Blanco 4m telescope on Cerro Tololo, at 2207 meters elevation in the Andes Mountains near La Serena, Chile. DECam has a 3deg^2 field of view, a new 5 lens optical corrector, and 250-micron thick fully-depleted red-sensitive CCDs. Photometric redshifts are obtained using 5 filters (g, r, i, z, and Y band) with central wavelengths 473, 642, 784, 926, and 1009 nm, respectively (Figure 3.5). The focal plane includes 62 of the $2k \times 4k$ CCDs that are used for imaging, and 12 smaller format 2k$\times$2k CCDs for guiding and focus/alignment (Figure 3.6). The five lens optical corrector is supported in a steel barrel and mounted to the prime focus cage with a hexapod that provides focus, lateral positioning, and tip/tilt capabilities. The shutter and filter changer are located between the 3rd and 4th lenses. The CCDs are cooled with a closed loop liquid nitrogen system and housed in a vacuum vessel mounted to the corrector barrel. The fifth lens of the corrector also serves as the window of the vessel. Figure 3.4 shows the DECam Reference design.

Three auxiliary detectors on the CTIO summit, supplied by DES, provide information for photometric calibration [149]. An All-Sky Radiometric Camera (RASICAM) is used to monitor the sky using the wavelength range 970 nm$< \lambda <$1250 nm. In this wavelength range, relatively warm clouds are easily distinguished from cold, clear skies. 'GPSMon' provides a cross-check of the amount of precipitable water vapor (PWV) in the atmosphere. And the Atmospheric Transmission Monitoring Camera (aTmCam) consists of a Paramount telescope mount and four small telescopes, each with a different narrowband filter, which monitors the brightness of suitable standard stars, thus providing the atmospheric transmission in wavelength regions dominated by the PWV and aerosol optical depth.

Generally the camera and telescope performance has been very good. Nonetheless, there has been a continuous effort to improve it [149]. After DES, the DECam will continue to be available as a Community Instrument on the Blanco Telescope for a long time.

3.4 Survey Strategy

The survey strategy for the DES consists of 10 separate tiling, each offset from the others by a significant fraction of the camera field of view, such that observations of individual astronomical sources are spread across the focal plane. A 'tiling' is a set of exposures, one in each of 5 filters. The g, r, i, and z-band exposures are 90 seconds duration. The Y-band exposures are 45 seconds duration through Y3. In Y4 the Y-band observations will be changed to 90 seconds.

First, an area of approximately 10% of the full survey solid angle of 5000deg^2 has been imaged to full DES depth. This allows an immediate verification of the data quality for the full survey. Second, the remaining areas of the survey will be imaged over successive seasons, building up approximately 2 layers of imaging in each band during each season. The DES tiling strategy involves creating a full layer of imaging in a single band covering the full 5000deg^2 area. This provides uniformity of coverage and control of systematic photometric errors via relative photometry on scales up to the survey size.

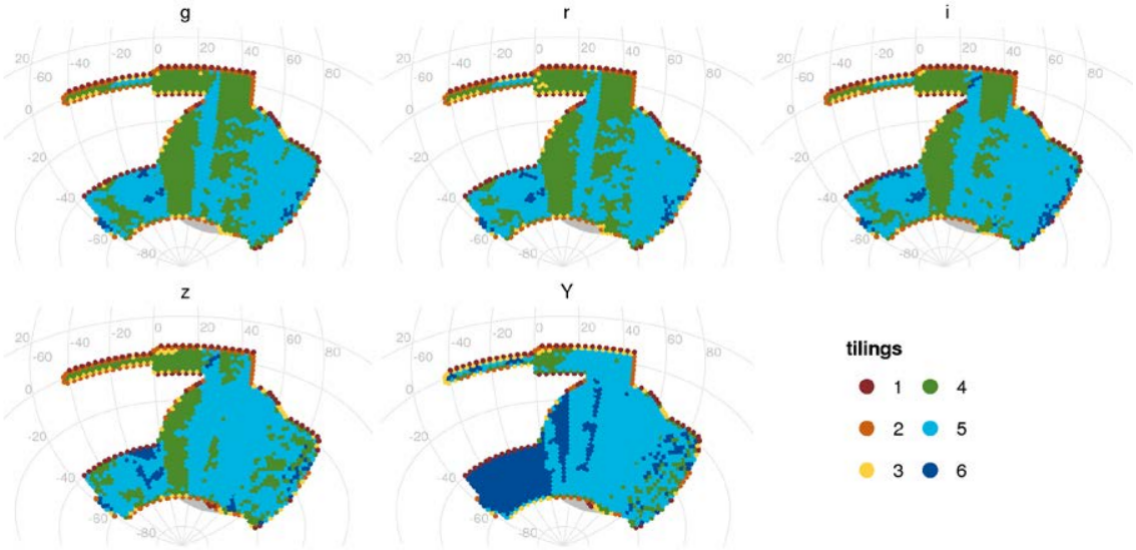


Figure 3.7: The completed Y1-Y3 survey fields in each filter. The colored dots represent the number of 'good' exposures [148].

For Y1 DES observed the SN fields plus a 2000deg^2 subset of the entire DES Wide Field (WF) footprint consisting of the first four tilings in each of the five filters. For Y2 they observed the SN fields plus the remainder of the DES wide field footprint up through the first four tilings/five filters that was not observed in Y1. For Y3 they observed the SN fields plus the entire DES wide field footprint up through the first six tilings/five filters that were not previously observed. They also had an interesting Target of Opportunity (TOO) arrangement with LIGO [158, 159] whereby 3 nights were added to the usual 105 nights. DES searches for the optical counterpart to the gravitational wave events triggered from the Advanced LIGO gravitational wave detectors. For Y4 the goal for wide field observing was to finish tilings 1-8 with priority given to those missed in previous seasons. The SN fields are successfully observed 20 to 28 times in a DES season. Y5 is nominally the final season for DES observing. They have to think carefully about how well the science goals of the survey can be achieved given the loss of time due to the unusually bad weather. If we are lucky, a Y6 will be granted and both the expected depth and area will be reached.

Despite having lost a lot of observational time due to the weather, DES has successfully observed 80% of an average first 3 seasons, and they do already have a reasonably uniformly covered wide field survey with at least four good exposures in all five filters. In Figure 3.7 the sky coverage in tilings at the end of Y3 is shown.

The survey field was designed to include complete overlap with the SZ cluster survey area covered by the South Pole Telescope (SPT) [160]. Also with part of SDSS [161] stripe 82 to provide tight constraints on the survey photometric calibration. Other spectroscopic surveys as 2dFGRS, VIMOS VLT Deep Survey, and DEEP2 will be used to calibrate empirical photo-z estimators, to measure photo-z error distributions, and to provide a sample of SN host galaxy redshifts. Besides, a follow-up spectroscopy of a subsample of $\sim 25\%$ of the SNe Ia on 8m-class telescopes is carried out.

3.5 Data Management (DESDM)

The DESDM system processes, calibrates and serves the DES data [162, 148]. The total data volumes are high ($\sim$ 2PB), and so considerable effort has gone into designing an automated processing and quality control system. A clear night of DECam observing is expected to produce about 300 science exposures together with approximately 60 bias and dome flat calibrations. The images are transferred by the NOAO Data Transport System (DTS) to NCSA/UIUC in Urbana-Champaign, Illinois, usually within 5 minutes after the moment that the shutter closes. Copies of the data are stored in La Serena and at the NOAO Science Archive in Tucson.

After each night of observing, the DESDM group processes the data for that night to generate initial catalogues, sky brightness estimates, point spread function (PSF), full-width at half maximum (FWHM) measurements and atmospheric opacity estimates for each exposure. This data is used to determine which exposures are of acceptable quality, and which need to be repeated [162, 148].

The 'Observing Tactician' (OBSTAC) processes the weather and previous data to select the highest priority fields to observe every night. Each night, the data is processed through a pipeline: image calibration (bias, flats, ...) in each band, astrometric calibration, image remapping and coaddition, point spread function modeling, and object detection and photometry (SExtractor). Supernovae are discovered by searching for temporal variations in brightness between SN template images and SN exposures. The images are processed by the DESDM difference imaging pipeline and object detection software to identify transient objects. These objects are then passed through machine learning algorithms to identify transients which are supernova candidates.

3.6 More than Dark Energy

Although optimized to measure and characterise dark energy, DES has already yielded both expected and unexpected discoveries beyond its primary goal for cosmological studies. Table 3.3 from [147] shows an inventory of measured and discovered objects by December 2015 (using SV, Y1 and Y2 data), as well as forecast for the complete survey. Highlights from DES early data include the discovery of 34 Trans Neptunian Objects, 17 dwarf satellites of the Milky Way, one published $z > 6$ quasar (and more confirmed) and two published superluminous supernovae (and more confirmed). This proofs that even if a project is optimised for dark energy searches, it can give outstanding results beyond cosmology.

Authors from reference [147] describe all non-cosmology science that can be done with DES:

- Science outside the original scope of DES: the solar system, the Milky Way, galaxy evolution, quasars (QSOs) , and transients.
- Astrophysics resulting from cosmology probes if the cosmology is assumed.

Objects	As December 2015	Expected from full 5yr DES
Galaxies with photo-z ($> 10\sigma$)	7M (SV), 100M (Y1+Y2)	300M
Galaxies with shapes	3M (SV), 80M (Y1+Y2)	200M
Galaxy clusters ($\lambda > 5$)	150K (Y1+Y2)	380K
SN Ia	1000	Thousands
Super-luminous SN	2 (published)+more confirmed+ +more candidates	15-20
New Milky Way companions	17	25
QSOs at $z > 6$	1 (published) + more confirmed + + more candidates	375
Lensed QSOs	2	100 ($i < 21$)
Stars ($> 10\sigma$)	2M (SV), 30M (Y1+Y2)	100M
Solar System:		
New Trans Neptunian Objects	32 in SN fields +2 un the WF	50 + more in WF
New Jupiter Trojans	19	
Main Belt Asteroids	300K (Y1+Y2)	
Kuiper Belt Objects		500-1000

Table 3.3: DES inventory (December 2015) [147]: different objects observed with DES over the SV, Y1 and Y2 seasons, and the expectation for the full 5 seasons.

3.7 Photometric Redshifts

Photometric redshifts are obtained from the multi-band photometry to produce a quasi-three dimensional survey. The calibration is done using about 15000 galaxies with spectroscopic redshifts available from other surveys (2dFGRS, VIMOS, VLT Deep Survey, DEEP2, Cosmic Evolution Survey field (COSMOS), OzDES and SDSS-I/II). A weighting method in a multi-dimensional color-magnitude space is applied to the spectroscopic sample in order to mimic the full DES photometric sample, which is on average significantly deeper than the calibration sample [163].

In several science analyses tomographic redshift bins are required. For these photo-z bins, the DES science requirements specify stringent limits on the differences in bias, scatter, and outlier fractions between the DES photometric sample and the spectroscopic calibration sample. The DES requirements for the scatter and the outliers are: $\sigma_{68} < 0.12$, $\text{out}_{2\sigma} < 0.1$ and $\text{out}_{3\sigma} < 0.015$ (where σ_{68} is defined from the 68% width about the median of $\Delta z = z_{\text{spec}} - z_{\text{phot}}$)).

The performance of several photo-z methods applied to SV data was evaluated in [164], using most of the relevant photo-z codes available. The best methods, Artificial Neural Networks or Random Forests, yielded a $\sigma_{68} = 0.08$. The results from most of the codes, including template fitting methods, comfortably meet the DES requirements on photo-z performance, therefore providing an excellent precedent for future DES data sets. This is expected to be improved with infrared photometry from VISTA Hemisphere Survey (VHS) survey, and studying the effects of completeness of the spectroscopic calibration samples.

The photo-z used in this thesis are:

- BPZ (Bayesian Photometric Redshifts) from [165] is a template-based method that returns the whole probability density distribution $P(z|m_i)$ that the galaxy is at redshift z when its magnitudes in each band are m_i , and also a single photo-z value computed as the maximum of the distribution.
- TPZ [166] is a machine learning parallel algorithm that uses prediction trees and random forest techniques to produce both robust photometric redshift pdfs and ancillary information for a galaxy sample.

3.8 The DES SV Benchmark Data Sample

In this thesis I perform measurements of the density contrast distribution and its moments on the DES Science Verification (SV) photometric sample (Figure 3.8) [4]. The DES Science Verification observations took place between 2012 and 2013 and provided data of over 250 deg^2 at close to the nominal depth of DES. From this sample we make selection cuts in order to recover the LSS Benchmark sample [142]. This minimizes the possible systematic effects and ensures the completeness of the sample [142]. This sample is focused on the SPT-E field with $60^\circ < \mathrm{RA} < 95^\circ$, and $-60^\circ < \mathrm{Dec} < -40^\circ$ considering only objects such that $18 < i < 22.5$ where i is `MAG_AUTO` as measured by `SExtractor` [167] in the i-band. We also perform the same color cuts found in [142]. The star-galaxy separation is performed by selecting objects such that `WAVG_SPREAD_MODEL` > 0.003. The total area considered for our study is then 116.2 deg^2 with approximately 2.3 million objects and a number density $n_g = 5.6$ arcmin^{-2}. Several photo-z estimations are available for these data. We will focus on the TPZ catalog [166].

This sample is ideal for our measurements since it is one of the first public Stage III imaging datasets, and the methodology presented here can be easily extended to the full DES area or other imaging datasets.

3.9 Simulations

The purpose of cosmological simulations is to model the growth of structures in the Universe. These simulations play a very significant role in cosmology because they can be considered as an 'experiment' to verify theories of the origin and evolution of the Universe. This was the idea behind the Blind Cosmology Challenge (BCC), where DES collaborators were provided with different simulations with unknown underlying cosmological models. In this thesis I use Buzzard-v1.0 from the BCC simulation suite, produced for DES [168]. These catalogues have previously been used for several DES studies (see e.g. [169, 170, 171, 172, 66]). The underlying N-body simulation is based on three cosmological boxes, a 1050 Mpc/h box with 14003 particles, a 2600 Mpc/h box with 20483 particles and a 4000 Mpc/h box with 20483 particles, which are combined along the line of sight producing a light cone reaching DES full depth. The simulation covers a quarter of the sky, with redshift

[4] This sample is available at `https://opensource.ncsa.illinois.edu/confluence/display/DESDM/DES+SVA1+Data+Products`

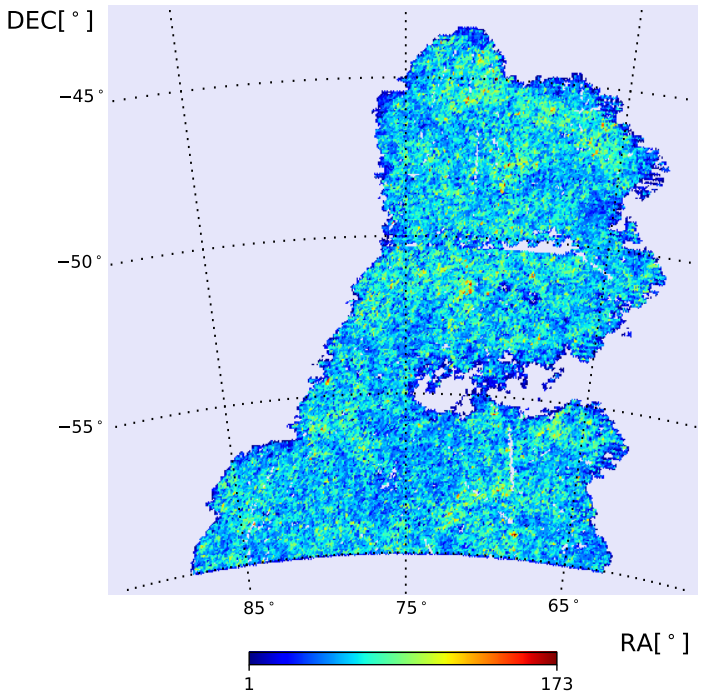
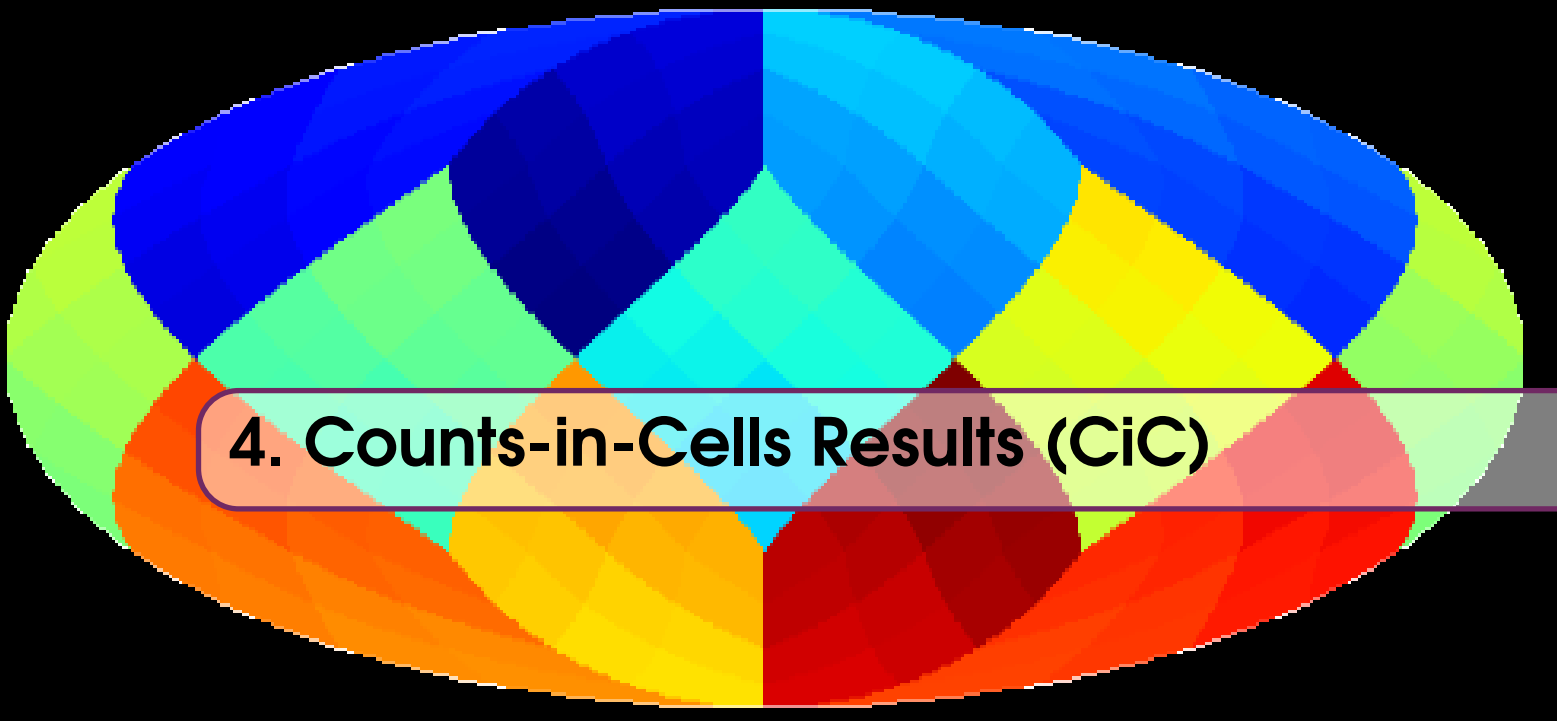


Figure 3.8: Footprint of the DES SV benchmark sample. Several cuts have been performed in the data in order to minimize systematic effects. This area contains approximately 2.3 million objects that we use in the CiC study (4).

z, between 0 and 2 and contains 980 million galaxies in the lightcone. The parameters of the simulation are from a ΛCDM model with $\Omega_m = 0.286$, $\Omega_\Lambda = 0.714$, $\Omega_b = 0.047$, $n_s = 0.96$, and $\sigma_8 = 0.82$.

Another DES simulation is the MICE Grand Challenge Simulation [173, 174], run by the Institut de Ciències de l'Espai (ICE) at the Marenostrum Supercomuter in Barcelona. MICE is an N-body simulation with cosmological parameters of a flat ΛCDM model with $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_b = 0.044$, $n_s = 0.95$, and $\sigma_8 = 0.8$. The simulation covers an octant of the sky, with redshift z, between 0.1 and 1.5 and contains 55 million galaxies in the lightcone. The simulation has a comoving size $L_{box} = 3072h^{-1}Mpc$ and more than $8 \cdot 10^9$ particles [175].



4. Counts-in-Cells Results (CiC)

In recent years, photometric redshift galaxy surveys have arisen as powerful probes of the Large Scale Structure (LSS) of the Universe and dark energy. The scientific community is concentrating a huge effort in them, exemplified by the Sloan Digital Sky Survey (SDSS), the Dark Energy Survey (DES), and the future Large Synoptic Survey Telescope (LSST) and Euclid. The main advantage of these surveys is their ability to retrieve information from a vast number of objects, yielding unprecedented statistics for the different observables to study the large scale structure (LSS). Their biggest drawback is the lack of precision in the direction of the line-of-sight and the systematic effects that are associated with it. Well constrained systematic effects and robust observables are required in order to exploit the best performance of such surveys. In this context, simple observables like the galaxy number of counts have an important role in probing the robustness of the survey. In particular, the galaxy Counts-in-Cells (CiC) provides information about the LSS and gives an estimate of how different systematic effects can affect the measurements. The main advantage of this method compared to other known methods to study the LSS is that, although simple, it provides insight of higher-order statistical moments of the galaxy counts, despite not being very demanding computationally. Other methods that can provide information about higher-order statistics such as the three- or four-point correlation functions require a lot more of computational resources.

Understanding the relation between galaxies and matter is essential for the measurements of cosmological parameters [119]. The uncertainties in this relation strongly increase the errors in the dark energy equation of state or gravitational growth index from future galaxy surveys [23]. The bias is highly-degenerate with other cosmological parameters and an independent method to determine it allows to break these degeneracies and improve the overall sensitivity to the underlying cosmology.

In this chapter I present a method to obtain the bias using CiC. First, I describe the method in Section 4.1. Then, I present the results of the linear and non-linear bias measurements in MICE simulation in Section 4.2. Finally, I apply the method to real data from DES, the SV sample, in Section 4.3.

4.1 Bias computation from CiC

One of the most important applications of the CiC observable is the determination of the galaxy bias. Since we can relate the different order moments of the density contrast distribution with the bias parameters (eqs: 2.91, 2.95). Assuming a linear bias, b_g, we can relate the second-order moment of the density contrast distribution of galaxies and dark matter:

$$\langle \delta_g^2 \rangle = b_g^2 \langle \delta_m^2 \rangle \tag{4.1}$$

If we assume a linear growth of structure we can relate the bias at different redshifts, also with CiC we can compute the moments and the bias for different scales θ:

$$\langle \delta_g^2(z,\theta) \rangle = b_g^2(z,\theta) \langle \delta_m^2(z,\theta) \rangle = b_g^2(z,\theta) \frac{D^2(z)}{D^2(0)} \langle \delta_m^2(0,\theta) \rangle \equiv b_g^2(z,\theta) D_+^2(z) \langle \delta_m^2(0,\theta) \rangle \tag{4.2}$$

where $D(0)$ is the normalization of the growth factor at $a = 1$. We do not have access to the total matter density field δ_m. Thus, we use a simulation where the bias is known:

$$\frac{\langle \delta_{g\ DATA}^2(z,\theta) \rangle}{\langle \delta_{g\ SIM}^2(z,\theta) \rangle} = \frac{b_{g\ DATA}^2(z) D_{+\ DATA}^2(z) \langle \delta_{m\ DATA}^2(0,\theta) \rangle}{b_{g\ SIM}^2(z) D_{+\ SIM}^2(z) \langle \delta_{m\ SIM}^2(0,\theta) \rangle} \tag{4.3}$$

The dependence of $\langle \delta_m^2 \rangle$ is smooth with the cosmological parameters, especially at high redshift where non-linear collapse is absent. We assume that the simulation's cosmological parameters are sufficiently close to the real cosmological parameters and we take the linear approximation:

$$\frac{\langle \delta_{m\ DATA}^2(0,\theta) \rangle}{\langle \delta_{m\ SIM}^2(0,\theta) \rangle} \approx \frac{D_{DATA}^2(0)}{D_{SIM}^2(0)} \tag{4.4}$$

Then we have that:

$$b_{g\ DATA} = b_{g\ SIM}(z) \frac{D_{SIM}(z)}{D_{DATA}(z)} \sqrt{\frac{\langle \delta_{g\ DATA}^2(z,\theta) \rangle}{\langle \delta_{g\ SIM}^2(z,\theta) \rangle}} \tag{4.5}$$

To take into account the accuracy of this assumption, and the possible difference due to the cosmological parameters we compute the quantity:

$$\delta p = \frac{D_{DATA}^2(0)/D_{SIM}^2(0)}{\langle \delta_{m\ DATA}^2(0,\theta) \rangle / \langle \delta_{m\ SIM}^2(0,\theta) \rangle} \tag{4.6}$$

We do this by computing the theoretical correlation function of dark matter for Planck and the simulation cosmology and integrating them to obtain the variance in both cases. We see that for every redshift and scale $\delta p \sim 1$ within a 5% error.

Finally we have that the bias is:

$$b_{g\ DATA} = b_{g\ SIM}(z) \frac{D_{SIM}(z)}{D_{DATA}(z)} \sqrt{\frac{\langle \delta_{g\ DATA}^2(z,\theta) \rangle}{\langle \delta_{g\ SIM}^2(z,\theta) \rangle} \delta p} \tag{4.7}$$

We compute the bias, b_{SIM}, of the simulation minimizing a χ^2 fitting different theoretical two-point correlation functions with different bias factors. We can do this with high precision since the cosmology of the simulation is known. We assume a fiducial cosmology for the data growth factor. We checked that the error of assuming the cosmology is lower than the statistical error (Section 4.3.3). Since bias is very environment dependent, we make sure that the simulation we use to compare with has the same characteristics as the data. It is important that the simulation has the same mask, same magnitude cuts and same photo-z.

If we want to go to higher orders, we can use equation (2.95) to compute the higher order bias parameters:

$$\begin{aligned} b_2 &= S_3\frac{b^2}{3} - S_{3\ m}\frac{b}{3} \\ b_3 &= S_4\frac{b^3}{4} - S_{4m}\frac{b}{4} - 3S_{3\ m}b_2 - 3\frac{b_2^2}{b} \end{aligned} \tag{4.8}$$

In a photometric survey like DES we do not have high-precision information in the line-of-sight compared to spectroscopic surveys, but we can do CiC projected on the sphere. Given that we are working in the angular projection, we can use the theoretical projected values for a conic (2.82) to obtain the values for the skewness and kurtosis of dark matter. To take into account the photometric redshift and the projection effects due to the redshift bin width we will compute this conical values of the skewness and the kurtosis convolving them with the photometric redshift distributions.

Another way of computing the linear and non-linear bias parameters is using the moments of dark matter from a dark matter simulation. We developed the linear bias computation method before having access to a dark matter simulation, so we will keep this method as it just requires a galaxy simulation. For the non-linear bias computation we will use the skewness and the kurtosis of the MICE dark matter simulation used in [132]. The skewness and the kurtosis hardly depend on cosmology (eq: 2.81). Then, we can use the values of a dark matter simulation also for the real data analysis even if the cosmological parameters are different from the ones in the simulation.

Due to the level of precision achieved in our analysis we will not take non-linear and non-local bias contributions to the two-point correlation into account.

Throughout this chapter, we assume a fiducial flat ΛCDM+ν (one massive neutrino) cosmological model based on Planck 2013 + WMAP polarization + ACT/SPT + BAO, with parameters [176] $\omega_b = 0.0222$, $\omega_c = 0.119$, $\omega_\nu = 0.00064$, $h = 0.678$, $\tau = 0.0952$, $n_s = 0.961$ and $A_s = 2.21 \times 10^{-9}$ at a pivot scale $\bar{k} = 0.05\text{Mpc}^{-1}$ (yielding $\sigma_8 = 0.829$ at $z = 0$), where $h \equiv H_0/100\text{km s}^{-1}\text{Mpc}^{-1}$ and $\omega_i \equiv \Omega_i h^2$ for each species i.

Method

The CiC method, while deceptively simple, is an extremely powerful, model-independent probe of the density distribution of matter in a catalogue. By dividing the sky into cells (or pixels) of fixed size and area, and then counting the objects in each cell, we calculate the

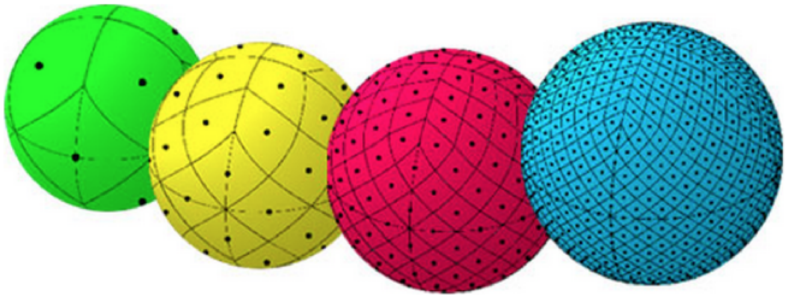


Figure 4.1: Healpix pixelization for different pixel sizes from 12 ($nside = 2^0$) pixels on the left to 768 ($nside = 2^3$) pixels on the right.

probability distribution $P(N)$ of finding objects in a cell.

The scheme of the code I developed is the following:

- Input
  - * Catalogue
  - * Mask
- Output
  - * Density contrast distribution
  - * Moments of the distribution

And the structure is as follows:

- Read data (galaxy catalogue ra dec z)
- Apply mask to the data.
- Pixelize the sphere and count the galaxies in each pixel
- Compute the density contrast in each pixel:

$$\delta = \frac{\rho}{\langle\rho\rangle} - 1 = \frac{N\mathrm{gal}_{\mathrm{pix}}/\mathrm{A}_{\mathrm{pix}}}{N\mathrm{gal}_{\mathrm{tot}}/\mathrm{A}_{\mathrm{tot}}} - 1 = \frac{N\mathrm{gal}_{\mathrm{pix}}\mathrm{A}_{\mathrm{tot}}}{N\mathrm{gal}_{\mathrm{tot}}\mathrm{A}_{\mathrm{pix}}} - 1 \tag{4.9}$$

  where N_{pix} is the number of pixels, Ngal is the number of galaxies and A is the area
- Fill an histogram with the density contrast
- Compute the moments of the density contrast distribution
- Compute the errors of the moments bootstraping (random sampling with replacement)

In most previous studies, these cells were spheres with different aperture radii [49], [88]. We perform our measurements by dividing the sphere into `HEALpix` pixels [177] (Fig: 4.1). `HEALpix` is an acronym for Hierarchical Equal Area isoLatitude Pixelation of a sphere and was originally developed by NASA's Jet Propulsion Laboratory to map the Cosmic Microwave Background (CMB). As its name suggests It is a pixelization of the sphere in pixels of the same area. The number of pixels is $N_{\mathrm{pix}} = 12 \cdot nside^2$, where the total number of pixels is controlled by the parameter *nside* which can only be a power of 2. Using this scheme guarantees that every pixel has the same area, regardless of its position

on the sphere. At the same time, the mask for DES - SV and other related DES products are stored using this scheme, so we avoid any potential boundary effects when using it. One downside to adopting `HEALpix` is that the number of pixels in which we divide the sphere cannot be chosen arbitrarily.

For our study we use pixels with *nside* ranging from 32 to 4096 (from 1.83° to 0.014°). The angular aperture, $\theta = \sqrt{Apix}$, is estimated as the square root of the pixel area. The angular scales mapped with these sizes match those analysed in [142]. According to equation (2.56) there is a dependence on the boundaries of the cell, and thus on the shape that we choose for the pixels. However, in [86], the authors estimate CiC for square cells of side l in a range $l = 0.03^o - 20^o$ and compare with the average correlation functions $\overline{w}_2(\theta)$. The agreement between the two estimates indicates that square cells give very similar results to circular cells when the sizes of the cells are scaled to $\theta = l/\sqrt{\pi}$. Using data from MICE in a thin redshift bin, $(0.95 < z < 1.05)$, we perform several tests to see that the concrete shape of the pixel, considering it is close to a regular polygon, does not affect the measured moments as regards to boundary effects. We compare rectangular pixels with `HEALpix` pixels. We divide the sphere into rectangular pixels taking n_{ra} parts in right ascension, and n_{ct} parts in $cos\theta$ where the number of pixels is $npix = n_{ra} \cdot n_{ct} = 12 \cdot nside^2$ (Figure 4.2). We take six different rectangle shapes named from 1 to 6. Pixels number 3 ($n_{ra} = 3nside$, $n_{ct} = 4nside$), 4 ($n_{ra} = 4nside$, $n_{ct} = 3nside$) and 6 ($n_{ra} = 6nside$, $n_{ct} = 2nside$) are close to being squares, but Pixels number 1 ($n_{ra} = 12nside$, $n_{ct} = 1nside$), 2 ($n_{ra} = 1nside$, $n_{ct} = 12nside$) and 5 ($n_{ra} = 2nside$, $n_{ct} = 6nside$) are far from being regular polygons. In Figure 4.3 we see that the difference is negligible for the more symmetrical pixels and higher for less symmetrical ones. When we compare square and `HEALpix` pixels we see that the measured moments are in perfect agreement. The advantage of using squares is that they can map any arbitrary scale. However, to deal with the DES masks and boundary effects it is more convenient to work with `HEALpix`.

The errors here, and throughout this chapter, are computed using the bootstrap method, random sampling with replacement. We take 100 random samples from the density contrast array of the same size as the length of the array. When the length of the array is shorter than 100 we take as many realizations as the length of the array. The error is the variance of those realizations.

When working with acquired observational data, the geometry of the survey becomes complicated. To deal with the boundary effects of an irregularly shaped area, we use the mask and degrade its resolution to match each of the pixel scales being used. However, degrading the mask (or increasing the scale) results in an increasing number of partially filled pixels. Only a fraction $r_A = A_{\text{filled}}/A_{\text{pixel}}$ remain completely inside the footprint. This means that, if we assign the same scale to all the pixels of a given *nside* value, some pixels will be effectively mapping a different scale. To solve this problem we can either require a minimum fraction of the pixel to be full, $r_A \geq X$, or we can compute the fraction of full pixels and perform CiC for that scale. We prefer to use the former since we consider that the scales where we perform the study appropriately map the variations of the density field in which we are interested. This approach also helps to avoid certain boundary effects. For small pixel sizes (similar to the size in the mask), given the large number of pixels we can

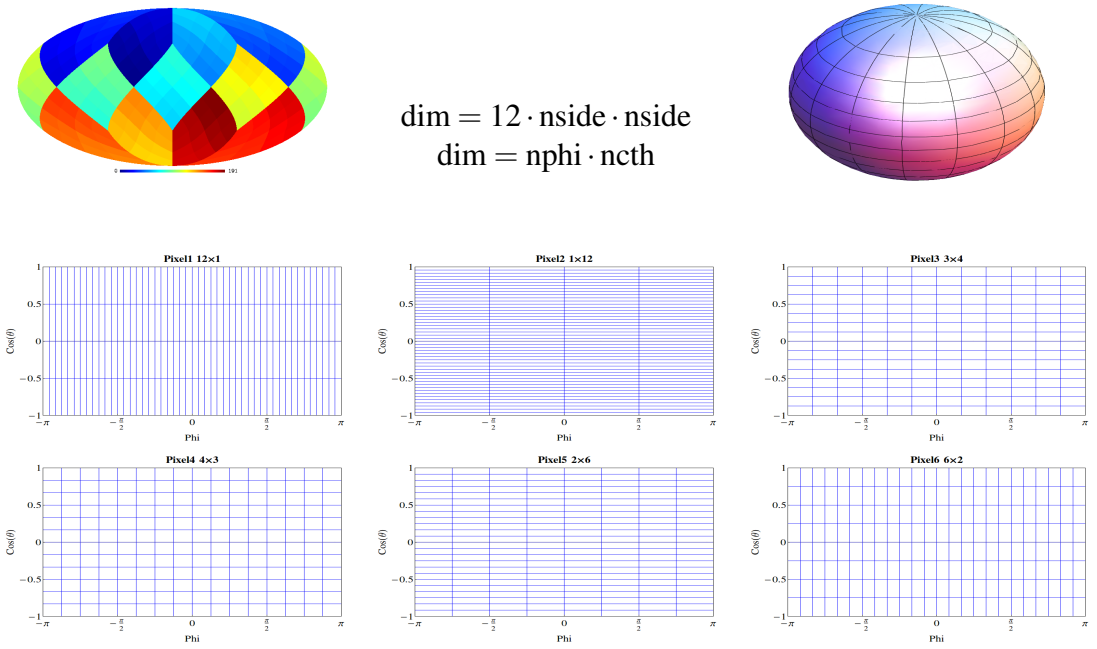


Figure 4.2: Different pixel shapes used to check the dependence of CiC with pixel shape. Pixels number 3 ($n_{ra} = 3nside$, $n_{ct} = 4nside$), 4 ($n_{ra} = 4nside$, $n_{ct} = 3nside$) and 6 ($n_{ra} = 6nside$, $n_{ct} = 2nside$) are closed to being squares, but Pixels number 1 ($n_{ra} = 12nside$, $n_{ct} = 1nside$), 2 ($n_{ra} = 1nside$, $n_{ct} = 12nside$) and 5 ($n_{ra} = 2nside$, $n_{ct} = 6nside$) are far from being regular polygons.

safely choose $r_A = 1$. For bigger pixels we try to find a compromise between the amount of area that we lose and the boundary effects. We set a threshold $r_A \geq 0.9$. In Figures 4.4 and 4.5 we show the area loss using data from MICE in the redshift bin $0.95 < z < 1.05$ with the SV mask for different thresholds in r_A. We see that all the pixels discarded (the blue ones) are in the boundaries of the mask. We see that if we choose pixels that are completely contained in the mask ($r_A = 1.0$) we lose a lot of area for smaller values of *nside*, however, very little area is lost for large values of *nside*. In Figure 4.6 we show the change in the moments for these different area cuts. It can be seen that the results are consistent for the different threshold values for r_A. We see that if we take all the pixels ($r_A \geq 0$) the difference in the moments is considerable in some cases, and we cannot take just all the pixels inside the mask ($r_A = 1$) because we run out of them for large scales.

In light of these results we are going to do CiC with healpix pixels and select only those pixels which are contained at least 90% inside the mask for *nside* ranging from 32 to 4096. When dealing with MICE simulation we take pixels entirely inside the octant to avoid boundary effects as the loss in area is very small.

4.2 Results for MICE simulation

Once we have checked the lack of dependency with cell shape, selected the optimal area cut and established the method for the computation of galaxy bias, we now turn to testing

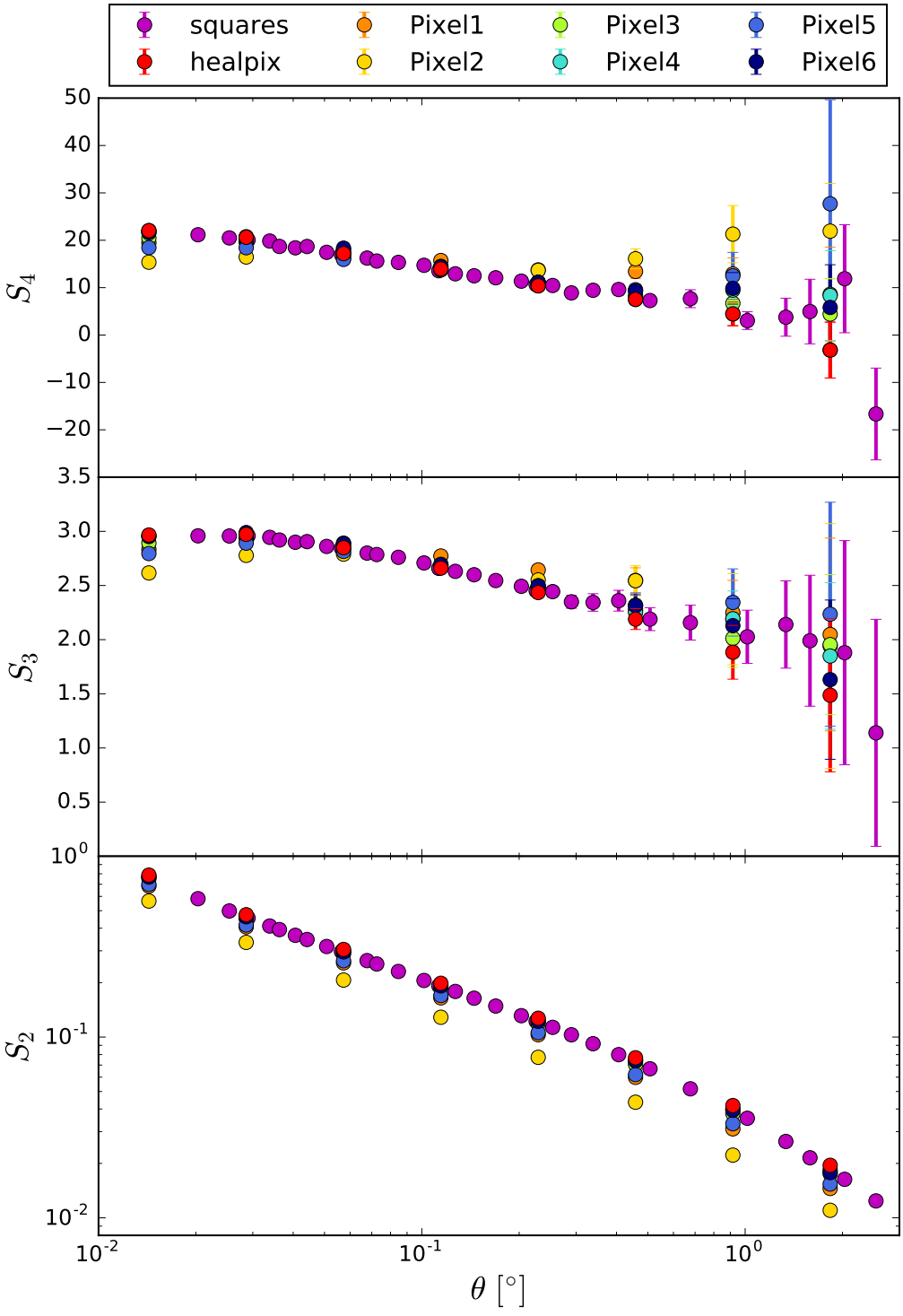


Figure 4.3: Moments of the density contrast distribution as a function of the cell scale, using data from MICE in the redshift slice $0.95 < z < 1.05$ for the different pixel shapes of Figure 4.2.

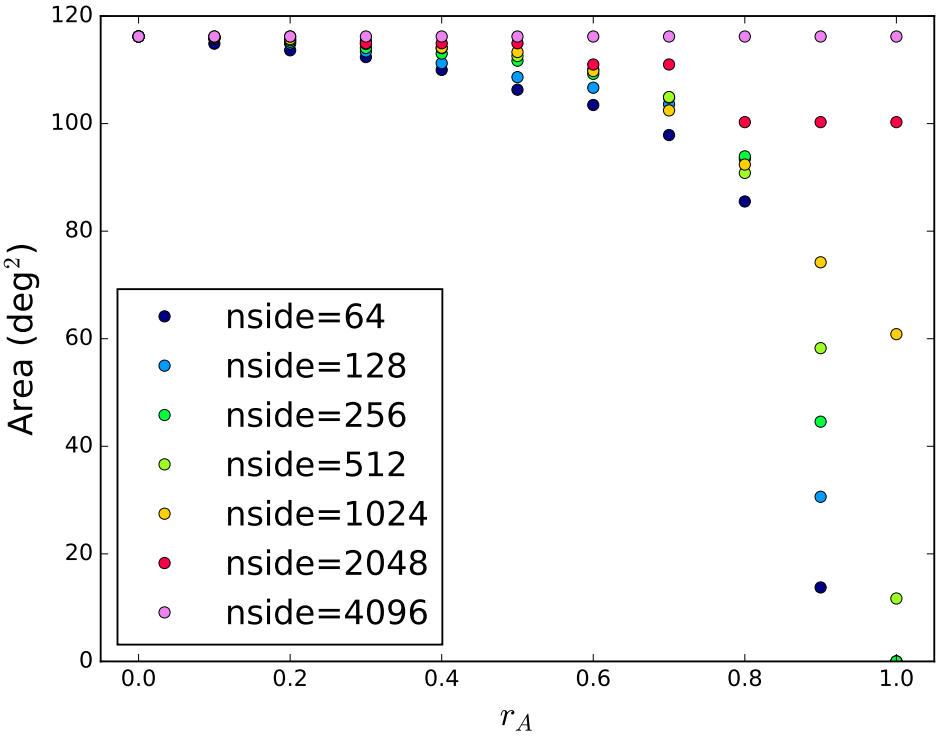


Figure 4.4: Area covered by different Healpix pixelisation resolutions, as a function of the minimum fraction of pixel coverage of said resolution with respect to the $nside = 4096$ footprint (larger pixels from lower nside will be partially filled at times). This test is done using the MICE simulation, considering the same footprint as the SV dataset.

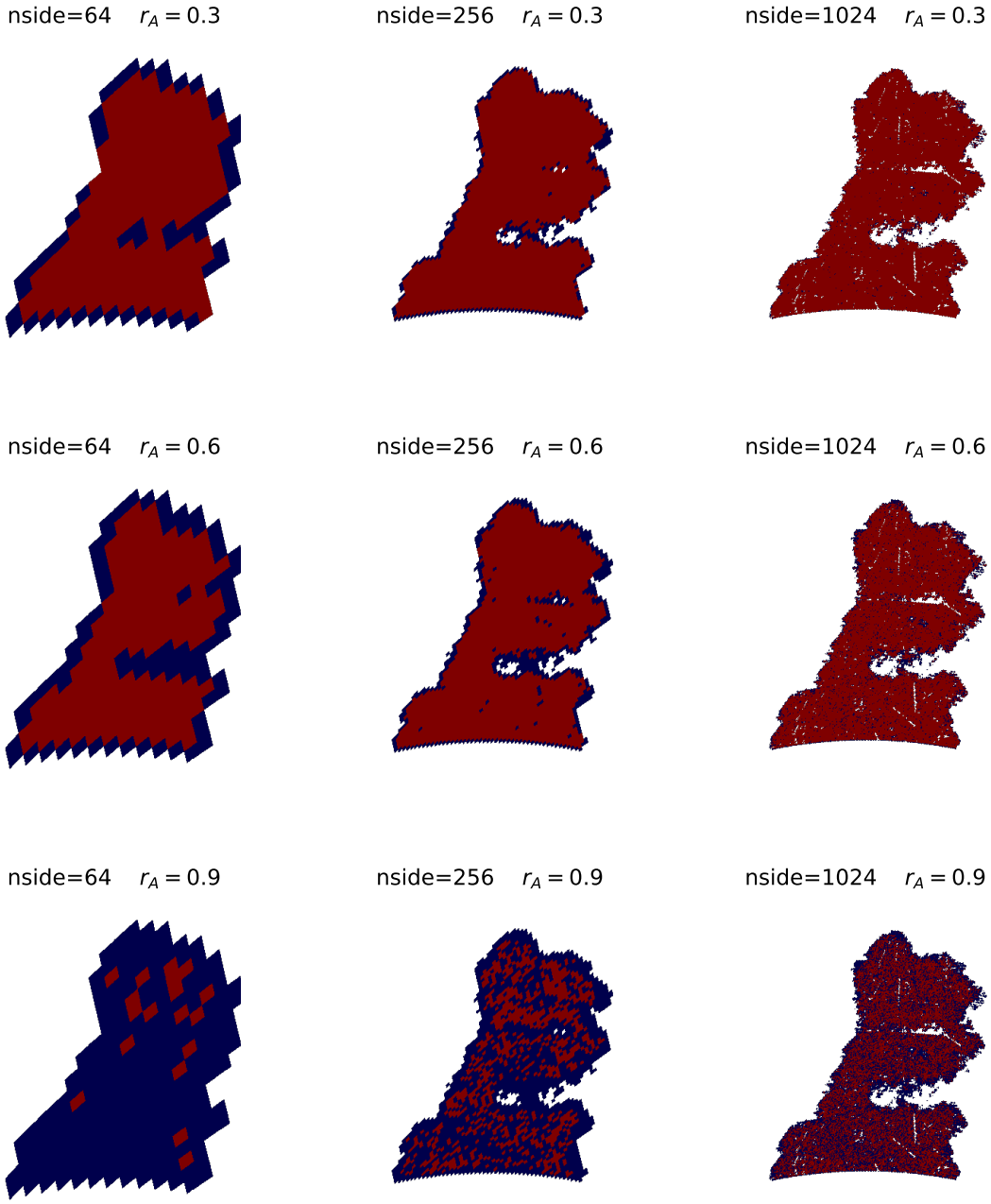


Figure 4.5: DES SV mask for different *nside* (64, 256, 1024) and different area cuts ($r_a = 0.3,\ 0.6,\ 0.9$). The pixels that we discard are blue and the ones that we keep are red. The bigger the pixel the more amount of data we lose for higher values of the area cut r_A.

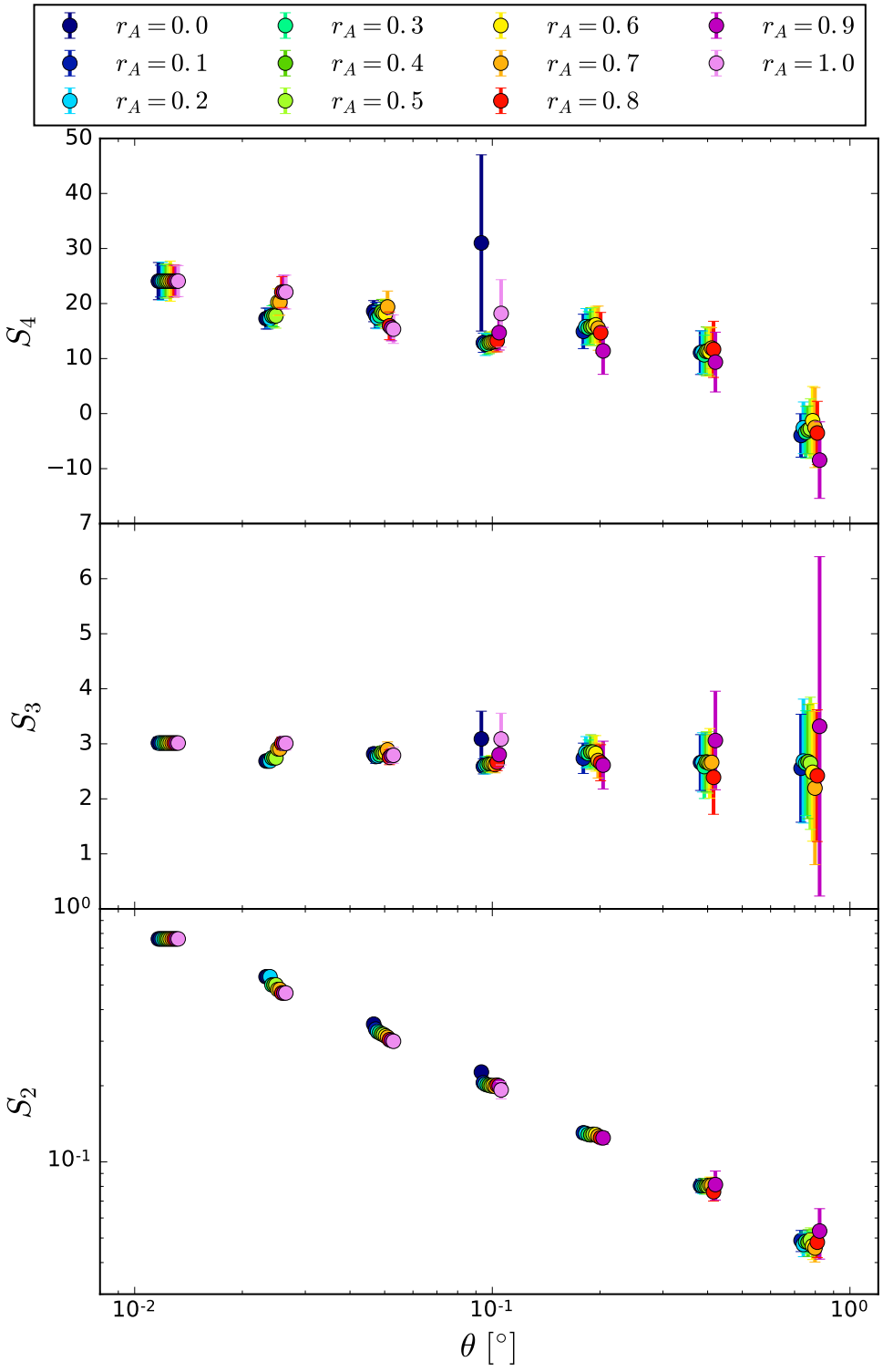


Figure 4.6: Moments of the density contrast distribution as a function of the cell scale, obtained from MICE ($0.95 < z < 1.05$) considering the SV footprint for different values of the fraction of the pixel that is contained in the mask, r_A.

the CiC method, described in Section 3 using the MICE simulation. We want to study the galaxy distribution throughout the history of the Universe. To do so, we divide the data in tomographic redshift bins and compute CiC in each of them. In order to have a galaxy population similar to that of the real data, we perform a magnitude cut in the i band $18 < i < 22.5$. First, we take thin redshift bins $\Delta z = 0.05$ in section 4.2.1, then we check how the redshift bin, the mask and the photometric redshift affect the moments of the density contrast distribution and bias computation in section 4.2.2. Since bias is very environment dependent we make sure that the simulation we use to compare with has the same characteristics as the data. It is important that the simulation has the same mask, same magnitude cuts and same photo-z.

4.2.1 Thin redshift bins

The simplest case, in order to compare with the theoretical prediction is to use thin redshift bins to avoid projection effects. To test our methodology in this scenario we divided MICE in many narrow top-hat redshift bins with $\Delta z = 0.05$. Figure 4.7 shows the moments in five of these redshift bins. We can observe that the moments follow the intuitively expected trend. The bigger the pixel, the smaller the value of the variance is. Since the pixel scales are bigger, we expect the pixels to be more homogeneous, and thus, have lower values for the variance. Figure 4.8 shows that we can fit the variance to a power-law, the slope gives us the tilt of the power-law power spectrum. The skewness and the kurtosis tend to a constant behaviour in the linear regime, above $\theta > 0.1^o$. In the non-linear regime they do not behave as expected (equations 2.77 and 2.80), mostly due to non-linearities and the low density of the simulation. But in the linear regime the values are consistent with equations (2.77) and (2.80) taking n from the power-law fit to the variance. Figure 4.9 shows higher order moments from S_3 up to S_7, in agreement with the same measurements in SDSS [123].

The evolution with redshift of the different moments is shown in Figure 4.10, we have taken the mean of the linear values ($\theta > 0.1^o$). They are constant for high redshifts and grow for low redshifts where non-linearities arise. The variance is lower for bigger scales (smaller *nside*) and starts growing below $z \sim 0.6$. The skewness and the kurtosis are also higher for lower scales (higher *nside*). For bigger scales ($nside = 64,\ 128$) they are constant down to $z \sim 0.3$, but for smaller scales ($nside = 1024,\ 2048$) they start growing below $z \sim 1.0$. This difference between different scales is because non-linearities affect further small scales. At large scales the Universe is homogeneous and we can neglect non-linearities.

CiC Check

To cross-check our results from CiC we use the correlation function. Equation (2.87) shows that we can obtain the variance integrating the angular two-point correlation function. The upper panel of Figure 4.11 shows the correlation function for $z = 1.0$ and the lower panel shows the comparison between the variance obtained from CiC and the one from the 2pcf.

The generalization for higher-order moments is straightforward (equation 2.86). However, we will not perform the theoretical predictions for these quantities since the access to

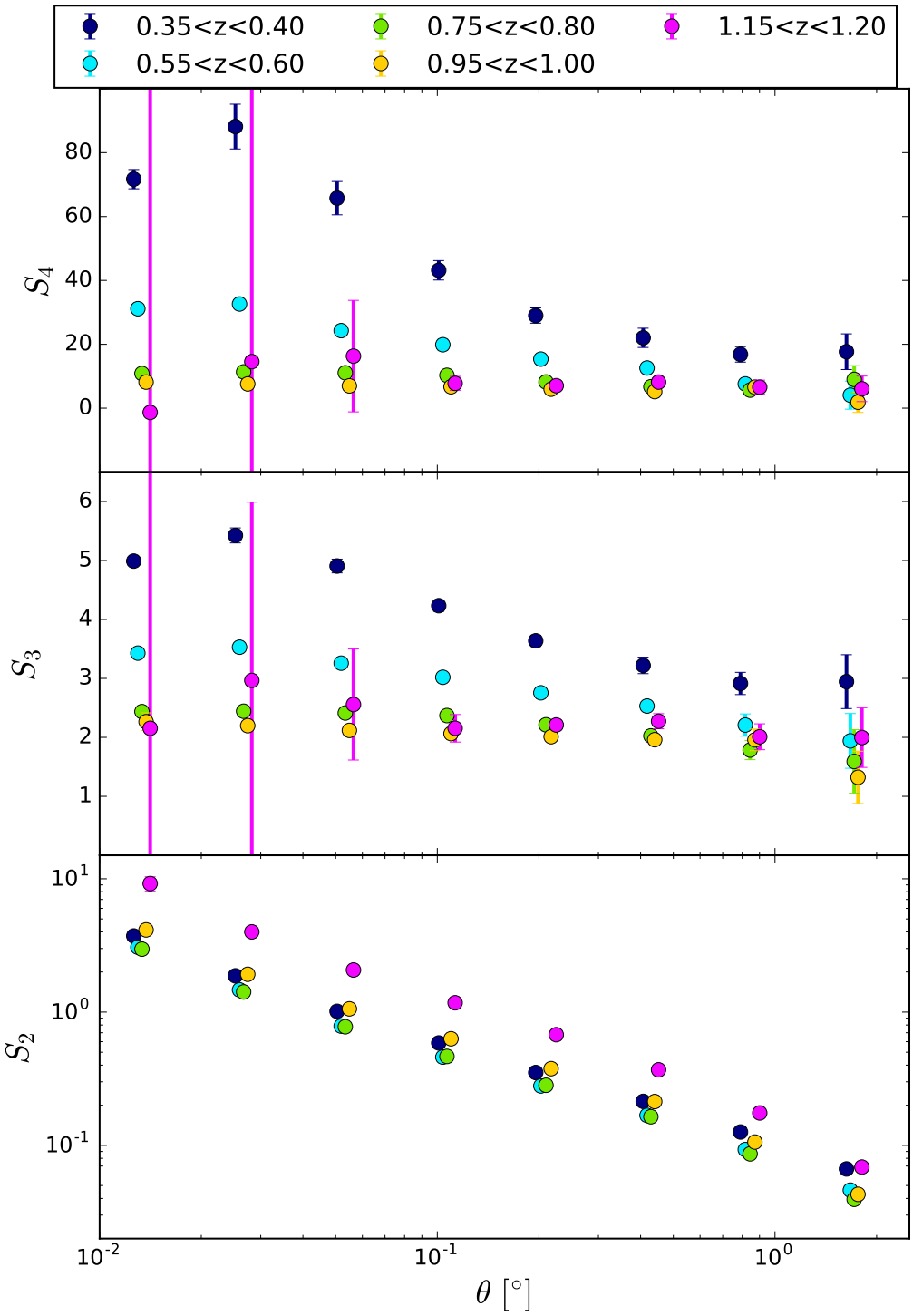


Figure 4.7: Moments of the density contrast distribution as a function of the cell scale, in the MICE simulation ($\Delta z = 0.05$) for five different redshift bins.

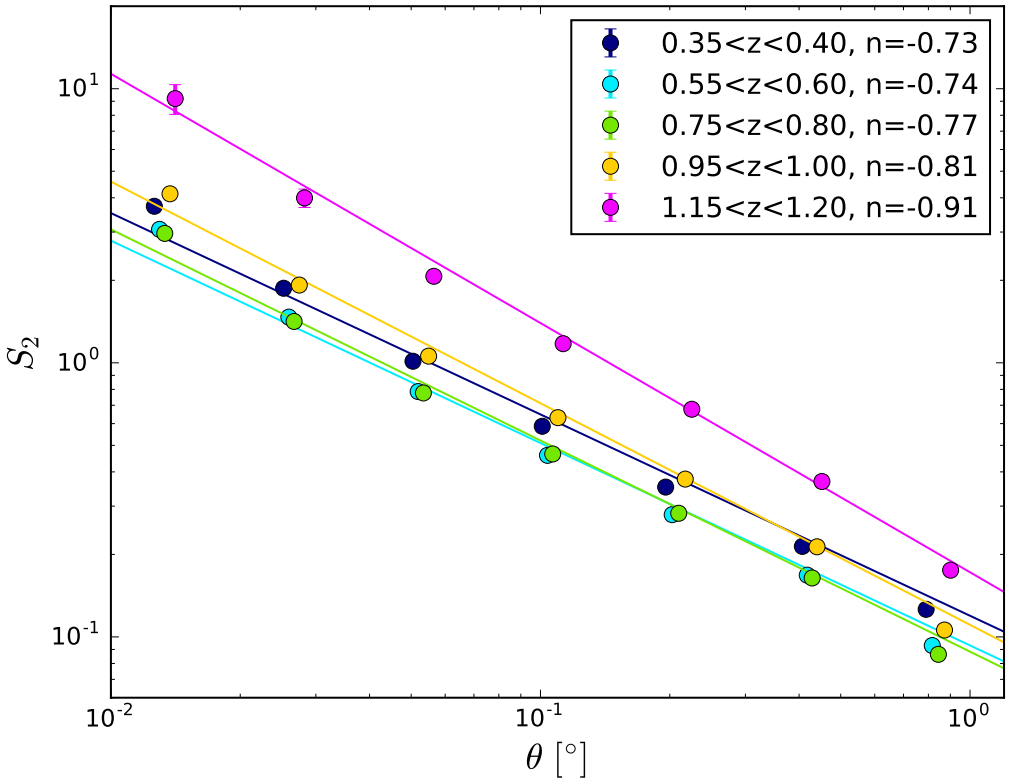


Figure 4.8: Power-law fit to the variance as a function of the scale in the MICE simulation ($\Delta z = 0.05$) for five different redshift bins.

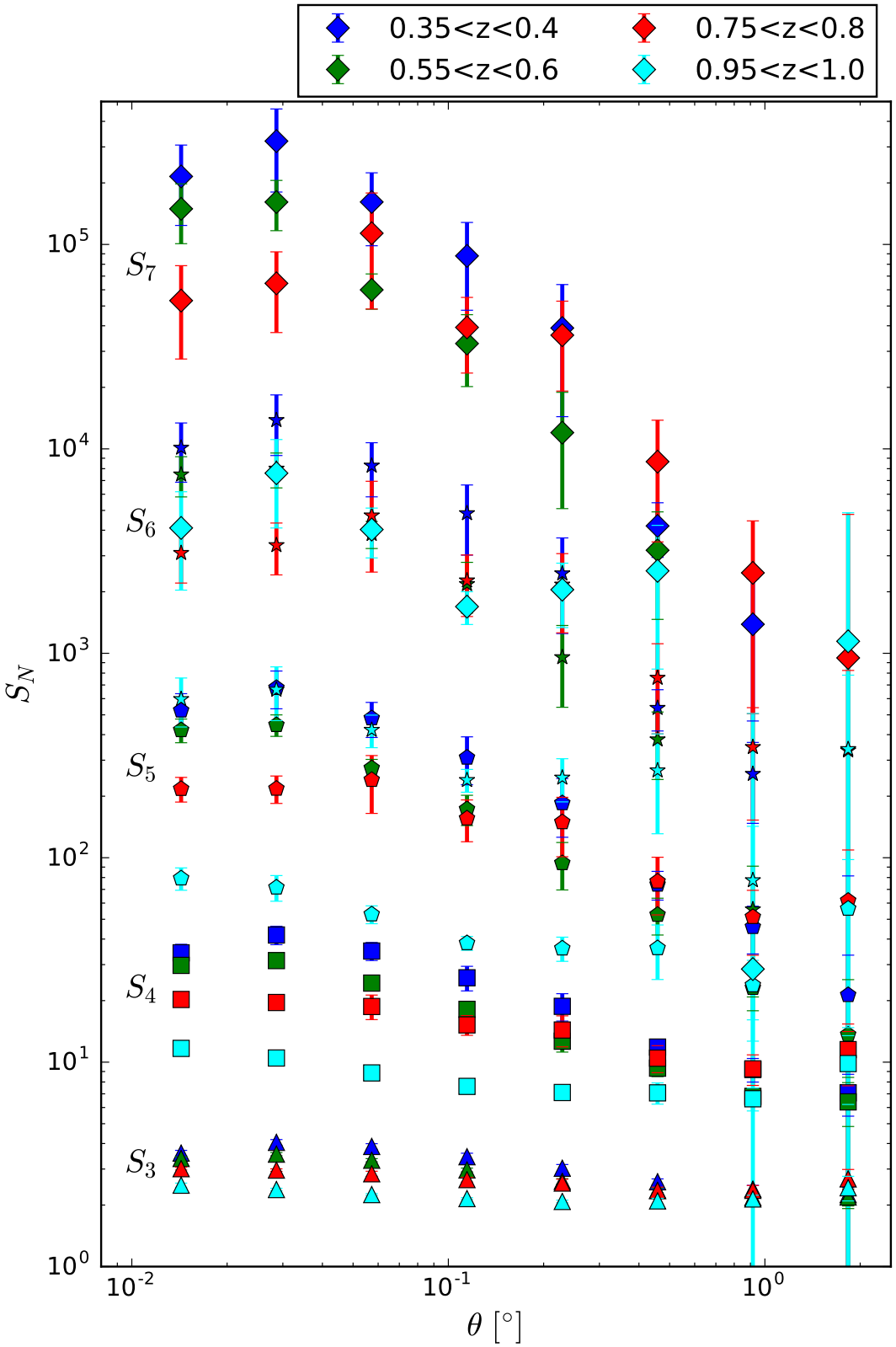


Figure 4.9: High-order moments ($3 < n < 7$) of the density contrast distribution as a function of the cell scale in the MICE simulation ($\Delta z = 0.05$) for four different redshift bins.

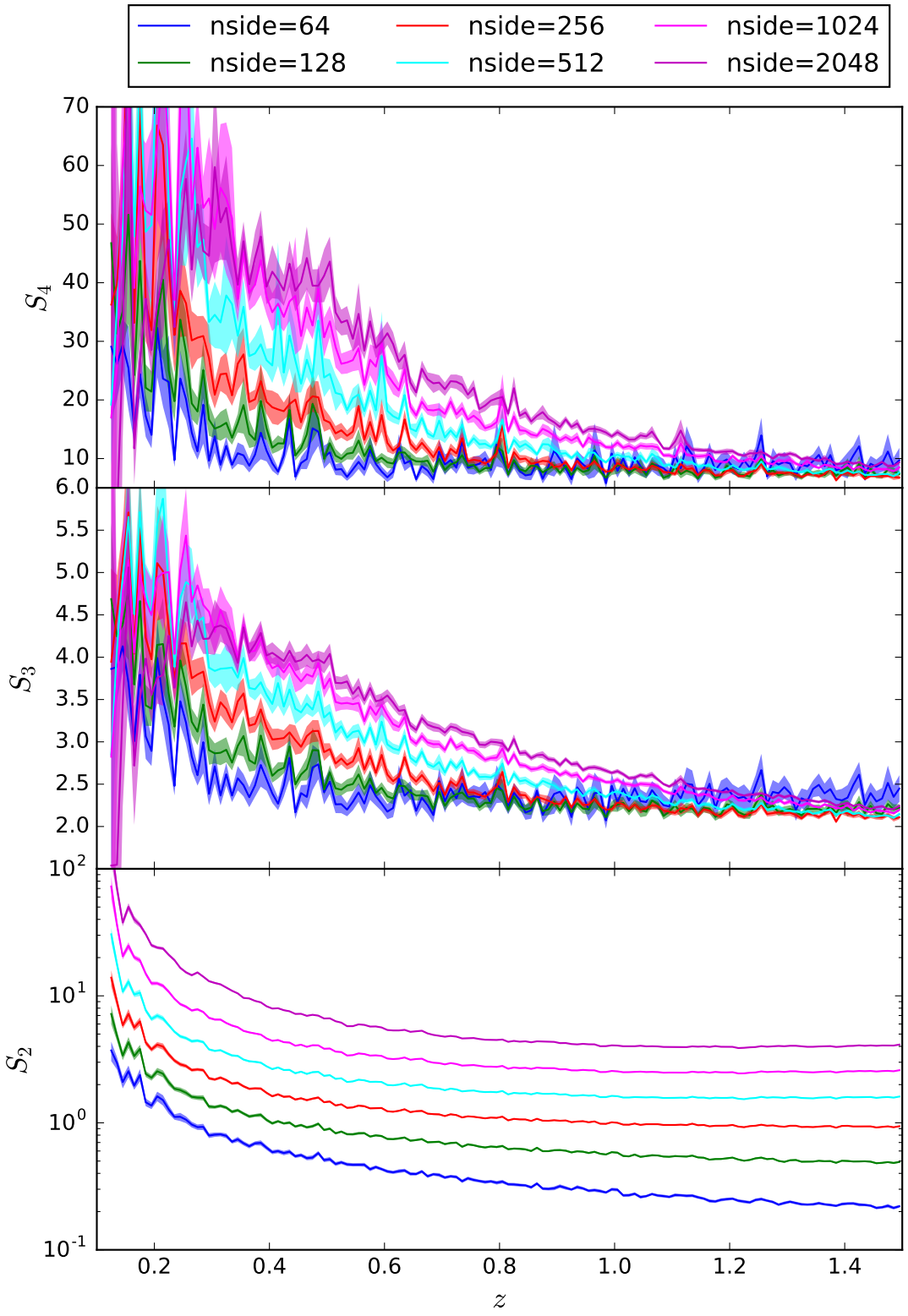


Figure 4.10: Evolution of the moments of the density contrast distribution with redshift in thin redshift bins ($\Delta z = 0.01$) in MICE simulation.

higher-order correlations is very demanding computationally. The strength of CiC relies on the fact that it can reveal properties about galaxy clustering that are not revealed by the two-point correlation function alone.

Linear bias

To obtain the linear bias of MICE and check the methodology, we use Buzzard simulation and apply equation (4.7) with MICE as the data and Buzzard as the simulation. The bias of the Buzzard simulation is obtained fitting theoretical two-point correlation functions $w(\theta)$ with different bias factors and minimizing the χ^2. In the upper panel of Figure 4.12 we see the bias as a function of the cell scale for five different redshift bins. Since, in this case, we are interested in the linear bias only, we just use the values in the shaded region. The scales below this region have a non-linear contribution, and scales above this region have large uncertainties due to lack of statistics. The linear scale changes with redshift, being higher for low redshifts due to non-linear gravitational collapse. In [142] they study the dependence of the linear scale with the redshift and find that it's always below $\theta = 0.1^o$. The bias is scale independent for the linear scales. In agreement with [110], who found that the local bias is consistent for scales larger than $R > 30 - 60$ Mpc/h. We are working in the linear scales between a few Mpc and up to ~ 45 Mpc. In the lower panel we find a good agreement between the bias of MICE, computed using CiC, and the one computed fitting theoretical two-point correlation functions with different bias factors and minimizing the χ^2 for the different redshift bins from $z = 0.1$ to $z = 1.3$.

MICE dark matter check

In the previous section we compared the bias obtained with CiC with the one obtained with the two-point correlation function. This just proves that both definitions of the bias are consistent. We would need the real dark matter density contrast for a real check. In order to do this, we compare the CiC measurement with the bias measurement in [132]. Figure 4.13 shows that our measurement is compatible with the one that the authors of [132] obtain comparing the angular two-point correlation function of galaxies with the one from dark matter. In previous sections we used the whole octant from MICE, and here we use the same data as [132], with $i < 22.5$ and $0 < (ra, dec) < 30$. We take thin redshift bins $\Delta z = 0.05$.

Non-linear bias

It is clear that non-linearities must be taken into account, they cannot be neglected at low redshifts or small scales. To obtain the values of b_2 and b_3 we use the linear bias measurements from the previous section (MICECATv2 and $\Delta z = 0.05$) and as explained in section 4.1 we compute the conic values from equation (2.82). In Figure 4.14 we compare this values with the ones obtained from the MICE dark matter simulation. The theoretical values are lower than the dark matter simulation ones and they do not vary with the redshift or with the scale. We conclude that it is better to use the dark matter values as they include the window function and can be modified to mimic the photometric redshift or any other features in the data. Whereas the conic theoretical value computed cannot mimic these features. Figure 4.15 shows the values of b_2 and b_3 computed with equation (4.8) with the

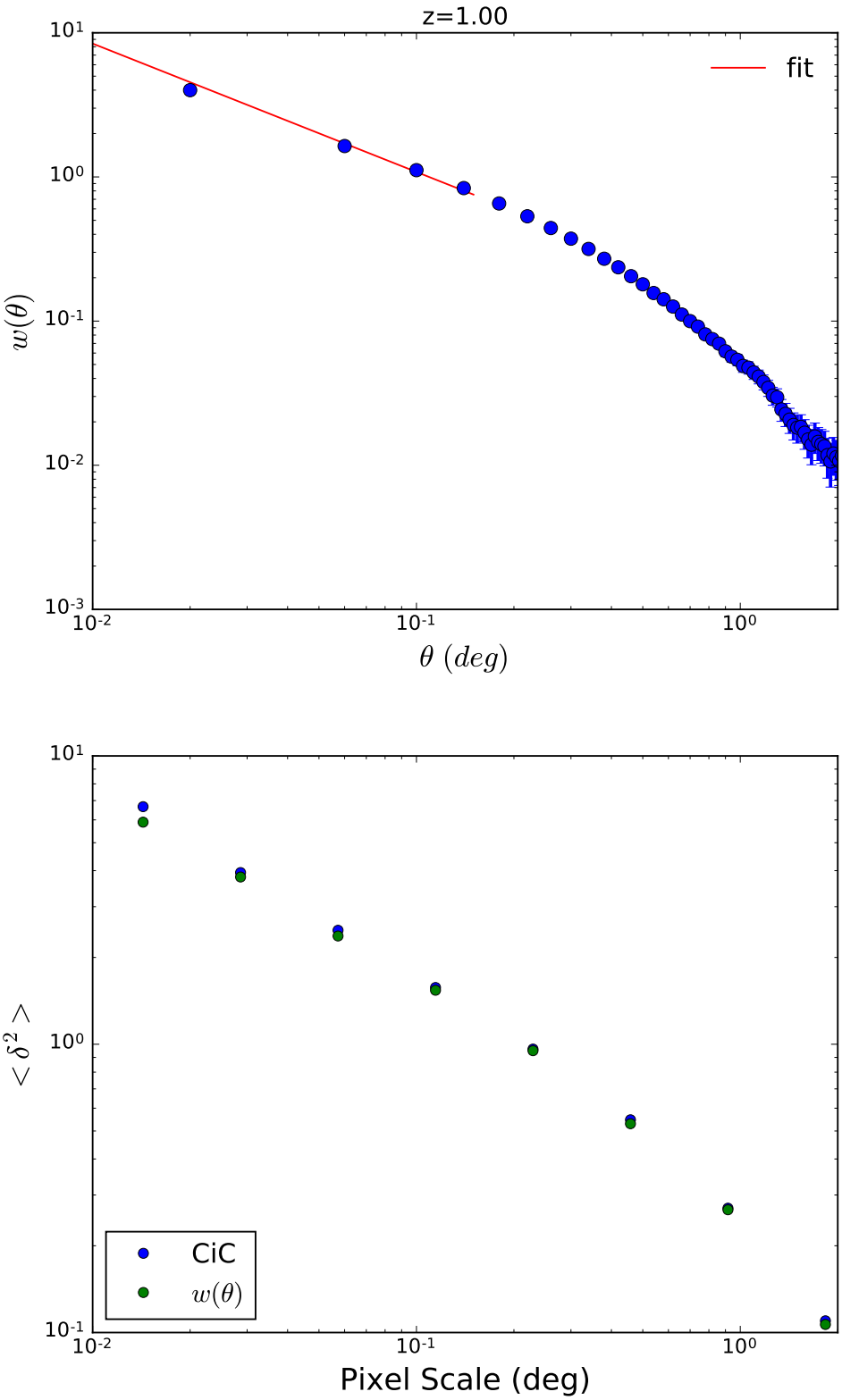


Figure 4.11: Upper panel: Angular two-point correlation function of $z = 1.0$. Lower panel: Variance as a function of the cell scale of $z = 1.0$ from CiC and from the integration of the upper angular 2pcf.

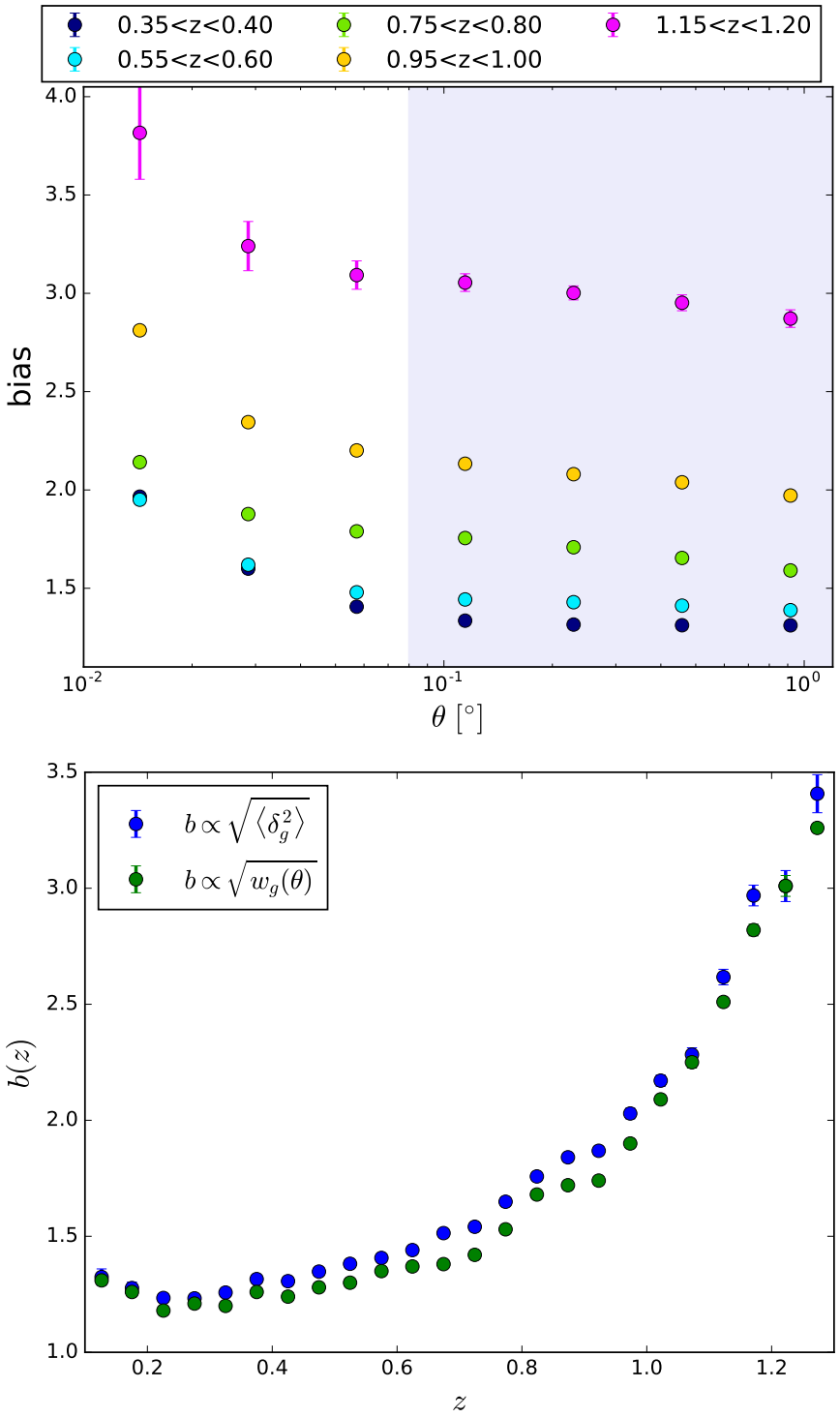


Figure 4.12: Upper panel: Bias obtained in MICE simulation ($\Delta z = 0.05$) from second order CiC as a function of the cell scale for five different redshift bins. The shaded region corresponds to the points that are in the linear regime $\theta > 0.1^o$. The linear bias is obtained with equation (4.7) with MICE as the data and Buzzard as the simulation. The bias of the Buzzard simulation is obtained fitting theoretical two-point correlation functions, $w(\theta)$, with different bias factors and minimizing a χ^2. Lower panel: Bias obtained from second order CiC compared with the one obtained fitting theoretical angular correlation functions with different bias factors. The points from CiC correspond to the average of the points in the shaded region of the upper panel.

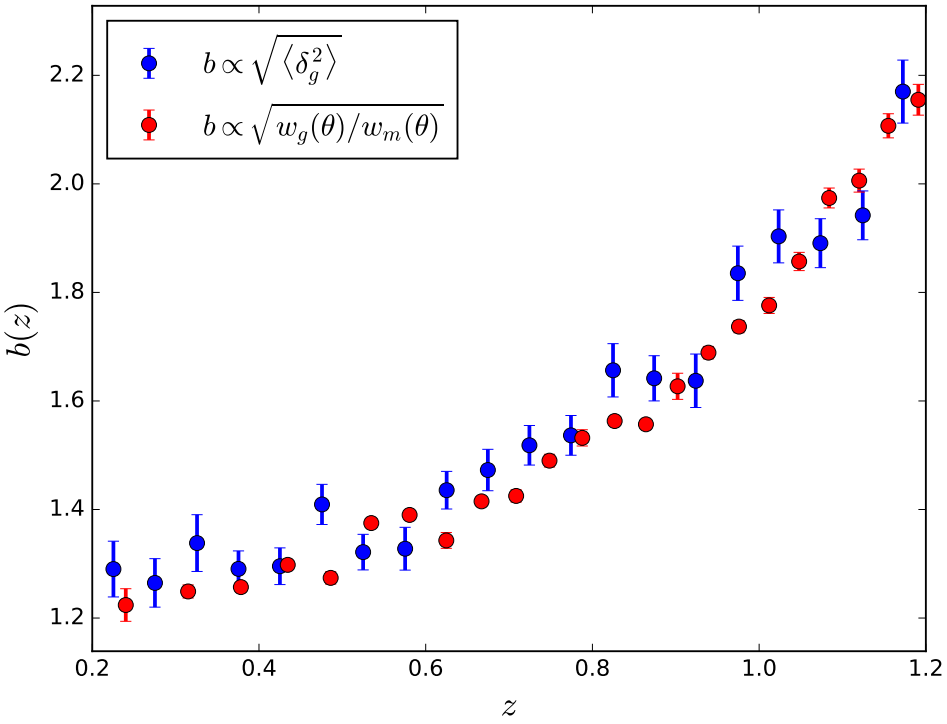


Figure 4.13: Bias obtained in MICE simulation in thin redshift bins ($\Delta z = 0.05$) from second order CiC compared to the one obtained by [132], comparing the angular two-point correlation function of galaxies with the one from dark matter. The linear bias is obtained with equation (4.7) with MICE as the data and Buzzard as the simulation. The bias of the Buzzard simulation is obtained fitting theoretical two-point correlation functions, $w(\theta)$, with different bias factors and minimizing a χ^2.

MICE dark matter simulation values.

We see in Figure 4.16 the comparison between the different orders of the non-linear bias expansion (equation 2.93). The figure shows the absolute value (sorted for the first-order term) of the different orders of the bias expansion for all the pixels in MICE $1.0 < z < 1.05$ and $nside = 64$ in the upper panel and $nside = 512$ in the lower panel. We see that at linear scales the higher order contributions are small but when we go to smaller scales they cannot be neglected. This is in agreement with [92].

4.2.2 Broad Redshift bins, mask and photo-z

Once we know the method works in ideal conditions for a simulation, we check how real data characteristics affect the measurements. We take wider redshift bins, apply a mask and a photometric redshift. All the plots corresponding to these cases can be found in the appendix A.

Different redshift bins

In photometric surveys like DES the redshift lacks the precision necessary to make these previous narrow redshift slices. If we performed such analysis, most points would be completely correlated in redshift. We will now check the method on MICE simulation using different redshift bin widths. We want to check how the redshift bin width affects the moments of the density contrast distribution. Figure A.1 shows the moments of the density contrast distribution as a function of the cell scale in MICE for different redshift bins ($\Delta z = 0.01, 0.05, 0.10, 0.15, 0.20, 0.25$) centred in $z_m = 1.0$. The variance is higher for narrower redshift bins, as wider redshift bins smear different redshifts. The Skewness and the Kurtosis do not change significantly with the bin width.

Broad Redshift bins

We will now check the method on MICE simulation using broad redshift bins of $\Delta z = 0.2$. We split our sample in 5 redshift bins: $z \in [0.2, 0.4], [0.4, 0.6], [0.6, 0.8], [0.8, 1.0], [1.0, 1.2]$. The same redshift bins used by [142], that we will use for DES data. Figure A.2 shows the moments of the density contrast distribution for the different redshift bins. We can observe that the moments here also follow the intuitively expected trend. The bigger the pixel value, the smaller the variance, and the skewness and the kurtosis are constant at higher redshifts, and vary at low redshifts due to non-linearities.

We compute the bias as we did before in the case of narrow bins. We use Buzzard simulation and apply equation (4.7) with MICE as the data and Buzzard as the simulation. The bias obtained for the different scales and redshifts is shown in the upper panel of Figure A.3. The points in the lower panel correspond to the average of the points in the shaded region of the upper panel, the linear region where the bias is scale independent. The CiC bias computation agrees with the bias of the simulation computed minimizing a χ^2 fitting different theoretical correlation functions with different bias. They also agree with the fit done to the bias of thin redshift bins ($\Delta z = 0.05$) in the lower panel of Figure 4.12.

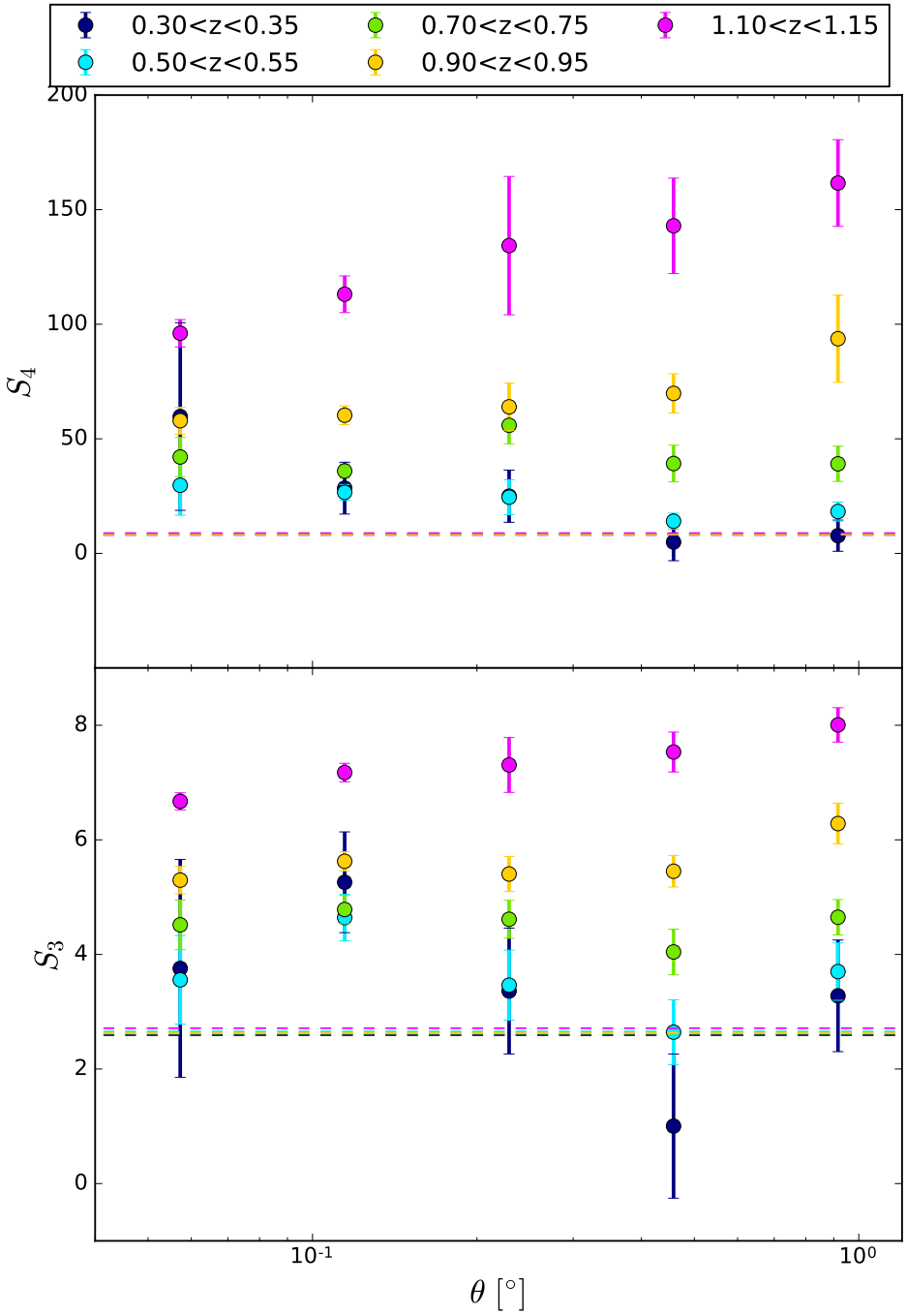


Figure 4.14: Comparison between the skewness and the kurtosis as a function of the scale obtained with equation (2.82) (dashed line) and the ones obtained doing CiC in the MICE dark matter simulation from [132] (points with error bars) for five different redshift bins.

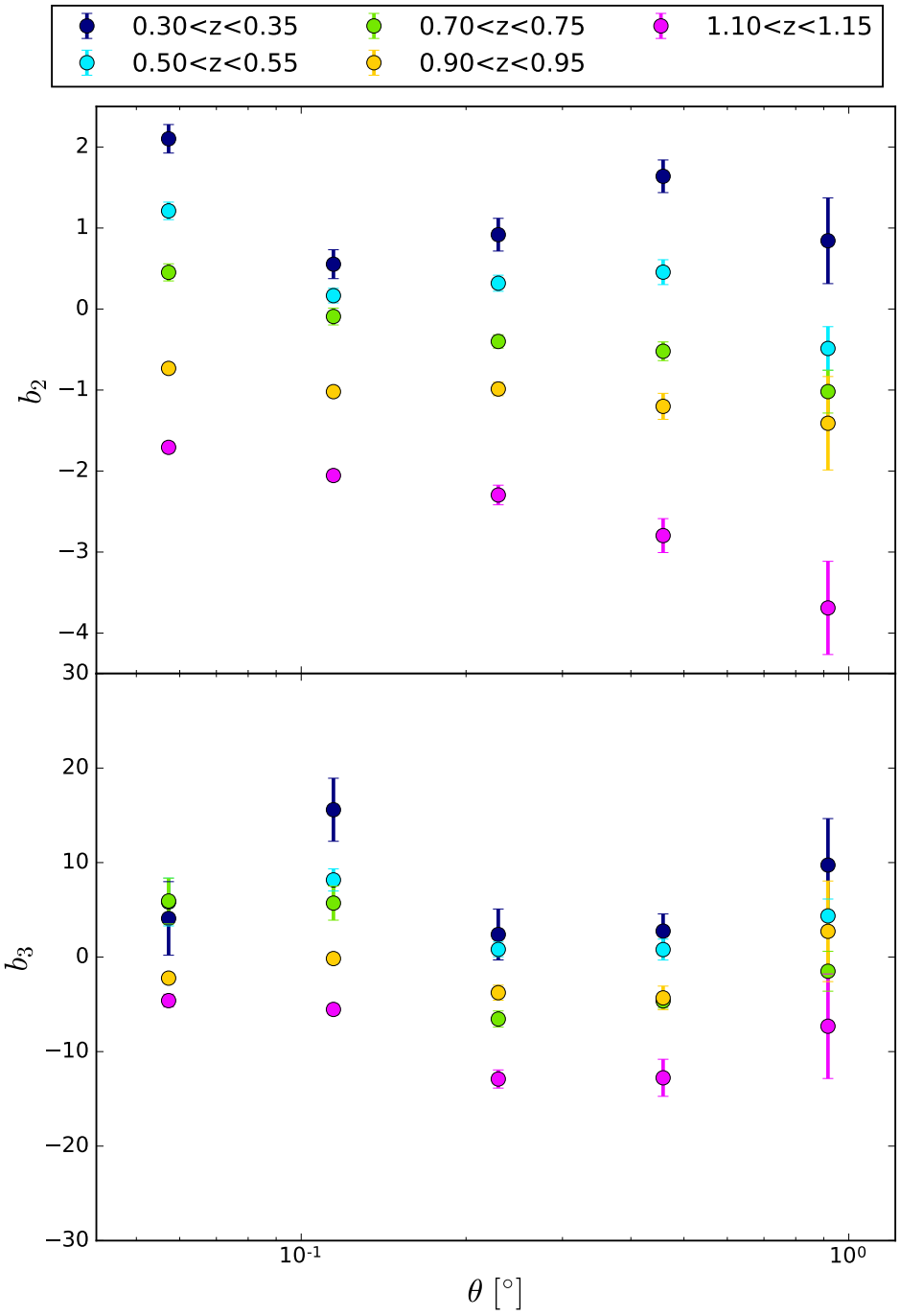


Figure 4.15: Non-linear bias parameters as a function of the cell scale obtained in the MICE simulation for five different thin redshift bins ($\Delta z = 0.05$). The values have been obtained from third and fourth order CiC with equation (4.8) with the MICE dark matter simulation values for the dark matter skewness and kurtosis.

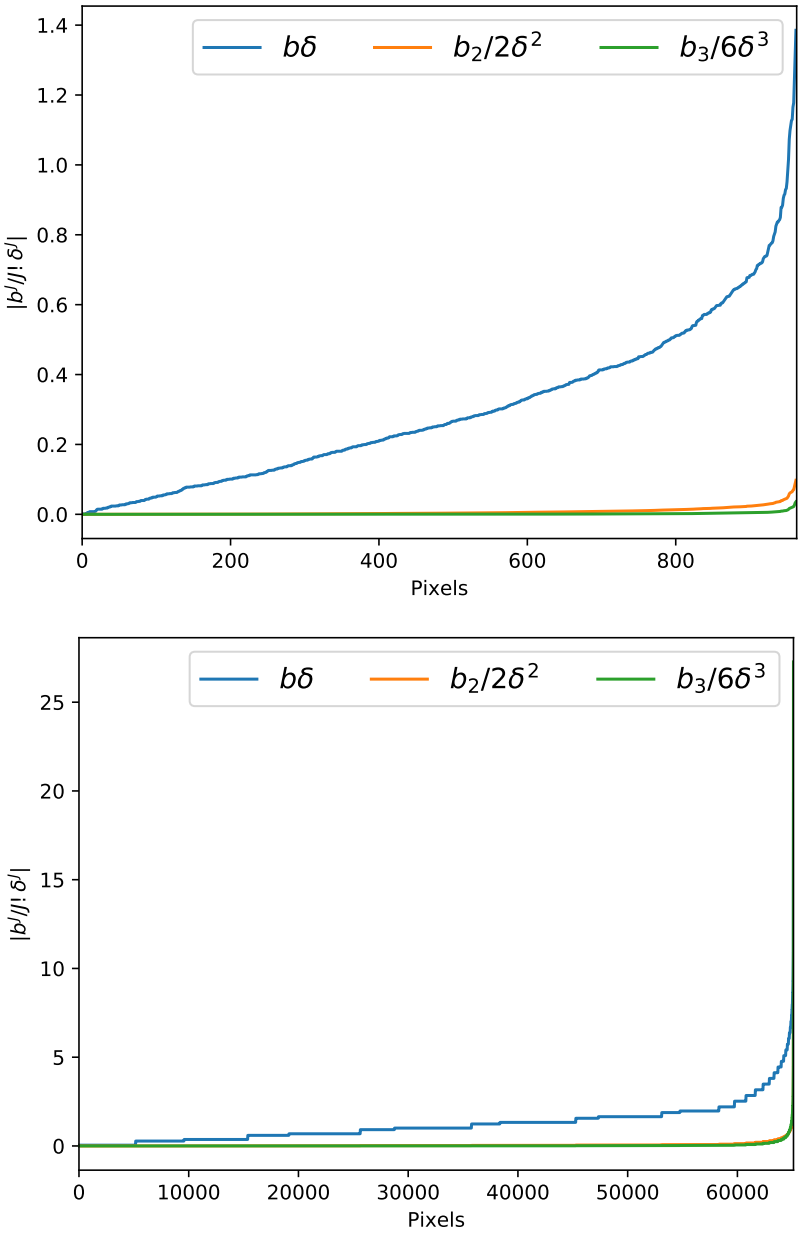


Figure 4.16: Absolute value of the different order terms in the non-linear bias expansion for the redshift bin $1.0 < z < 1.05$ in all the pixels in the mask (sorted for the first-order term). The upper panel shows a big scale, $nside = 64$, and the lower panel shows a smaller scale, $nside = 512$.

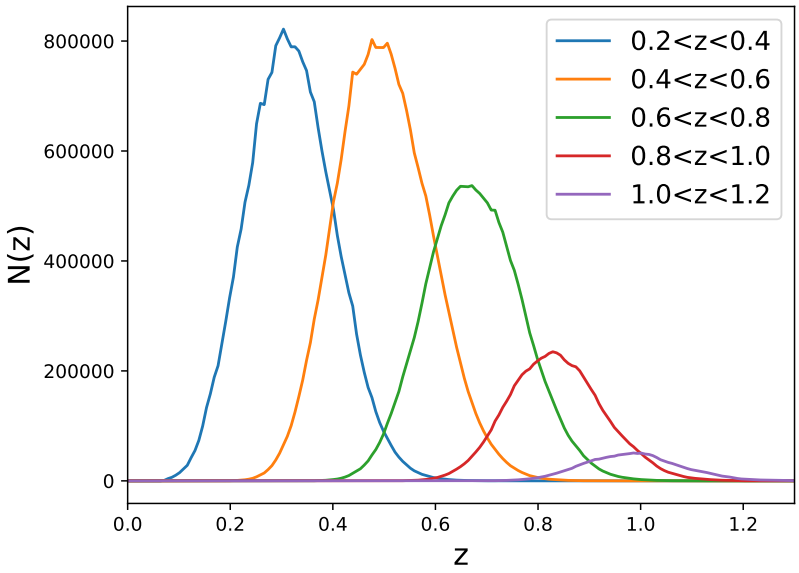


Figure 4.17: Redshift distribution of the MICE simulation with Gaussian photo-z $\sigma_z = 0.05(1+z)$.

We use the linear bias measurements and as explained in section 4.1 we obtain the values of b_2 and b_3 presented in Figure A.4, using the skewness and the kurtosis of the MICE dark matter simulation in the same redshift bins. They are consistent with constant values within the errors at linear scales ($\theta > 0.1°$), in agreement with [110], although they go to higher scales.

Photometric redshift

The uncertainty in the redshift distribution is an important systematic effect in photometric surveys like DES. We will now check the method on MICE simulation using a Gaussian photo-z smeared with $\sigma_z = 0.05(1+z)$, and broad photometric redshift bins $\Delta z = 0.2$ (Figure 4.17). This $sigma_z$ is of the same order as the real photometric redshifts used in the data, TPZ and BPZ. We split our sample again in 5 redshift bins. Figure 4.18 shows the moments of the density contrast distribution for the different redshift bins. The moments here also follow the intuitively expected trend but there is mixing in the variance between the different redshift bins due to the photo-z.

Once again we compute the bias using Buzzard simulation and applying equation (4.7) with MICE as the data and Buzzard as the simulation. The bias obtained for the different scales and redshifts is in the upper panel of Figure 4.19. The points in the lower panel correspond to the average of the points in the shaded region of the upper panel, the linear region where the bias is scale independent. We see in the lower panel that the CiC values agree with the bias of the MICE simulation computed minimizing a χ^2 fitting different theoretical correlation functions with different bias. We see that we are able to measure the linear bias with a photometric redshift. They are of the same order as the fit done to the bias of thin redshift bins ($\Delta z = 0.05$) in the lower panel of Figure 4.12. They do not exactly agree since the photometric redshifts change the redshift distribution and the bias

depends strongly on the redshift. Lower redshift bins include the bias of lower redshifts than the bin itself, and then are lower.

We use the linear bias measurements and as explained in section 4.1 we obtain the values of b_2 and b_3 presented in Figure 4.20, using the skewness and kurtosis of the MICE dark matter simulation in the same redshift bins and smeared also with a Gaussian photometric redshift. They are consistent with constant values in the linear scales.

MICE with DES-SV mask

In real data we have to deal with masks and boundary effects. We apply the SV mask to MICE simulation to see how this affects to the moments and bias computations, we take broad photometric redshift bins as in previous sections, $\Delta z = 0.2$. Figure A.5 shows the moments of the density contrast distribution for the different redshift bins and scales. The moments here also follow the intuitively expected trend but the errors are bigger, as the area is smaller than the octant of MICE and we see the influence of cosmic variance. The SV footpint area is 116.2 deg^2 and MICE is 44 times bigger, 5156.6 deg^2.

The upper panel of Figure A.6 shows the bias obtained using Buzzard simulation and applying equation (4.7) with MICE as the data and Buzzard as the simulation. As the area is small we take the average of 16 different realizations of the Buzzard simulation to avoid cosmic variance. The effect of cosmic variance in the simulations is shown in the systematic study in Section 4.3.3, and it is included in the errors. The points in the lower panel correspond to the average of the points in the shaded region of the upper panel, the linear region where the bias is scale independent. They agree with the bias of the simulation computed minimizing a χ^2 fitting different theoretical correlation functions with different bias. They also agree with the fit done to the bias of thin redshift bins ($\Delta z = 0.05$) in the lower panel of Figure 4.12.

We use the linear bias measurements and as explained in section 4.1 we obtain the values of b_2 and b_3 presented in Figure A.10 using the skewness and kurtosis of the MICE dark matter simulation in the same redshift bins. They are consistent with constant values in the linear scales but the errors are bigger than in previous cases due to the smaller area considered here.

MICE with DES-SV mask and Gaussian photo-z

In photometric surveys we have to deal with masks and boundary effects, but also with photometric redshifts. We apply a Gaussian photo-z and the SV mask to MICE simulation to see how this affects the moments and the bias computations. Figure A.8 shows the moments of the density contrast distribution for the different redshift bins. The moments here also follow the intuitively expected trend but the errors are bigger and the different redshifts mix due to the photo-z. We have lower statistics and cosmic variance due to the small area as in previous section, and smearing due to the photo-z.

The upper panel of Figure A.9 shows the bias for different redshifts and scales, obtained using Buzzard simulation and applying equation (4.7) with MICE as the data and Buzzard

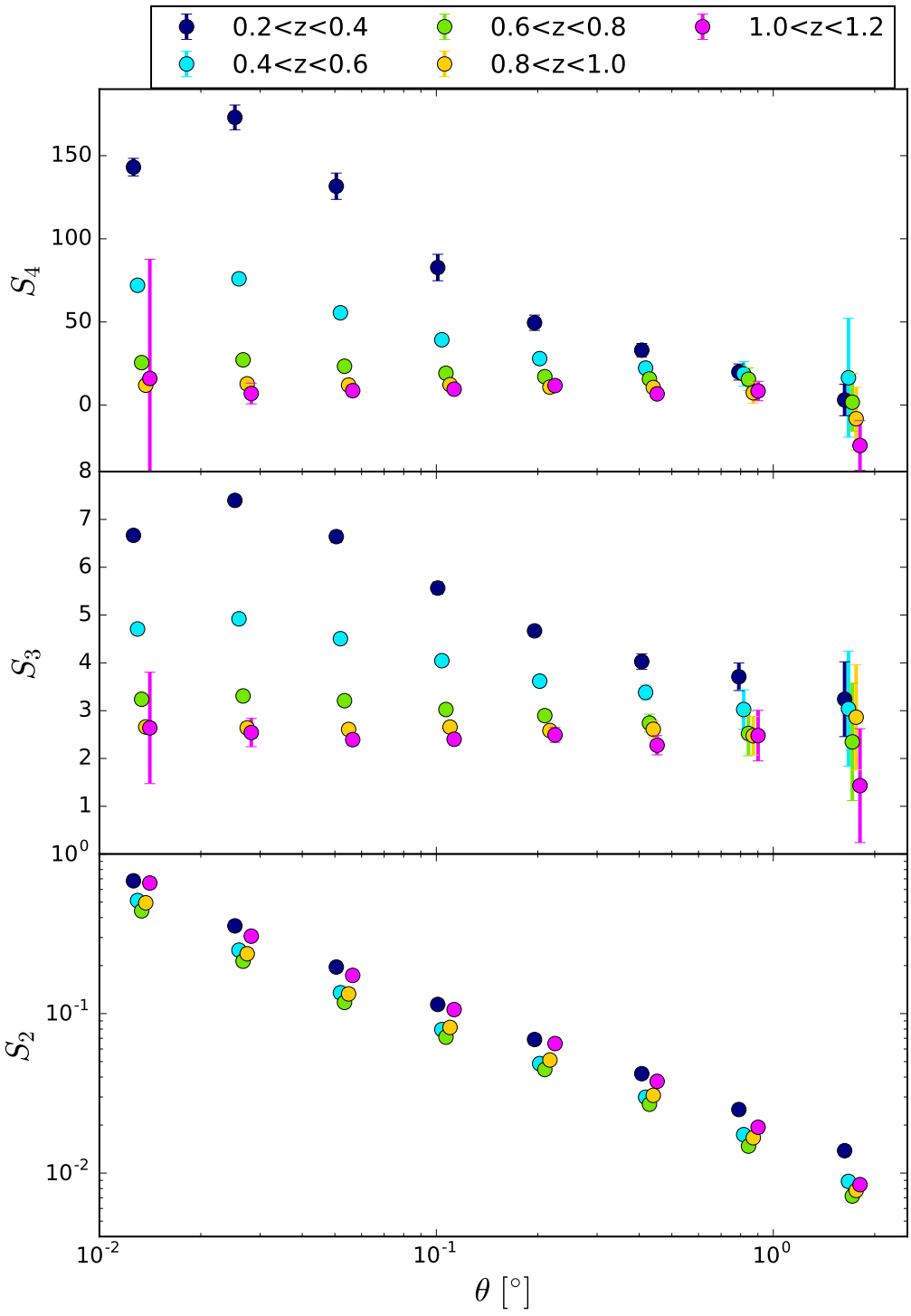


Figure 4.18: Moments of the density contrast distribution as a function of the cell scale in the MICE simulation with broad redshift and Gaussian photo-z redshift bins ($\Delta z = 0.2$ and $\sigma_z = 0.05(1+z)$) for different redshift bins.

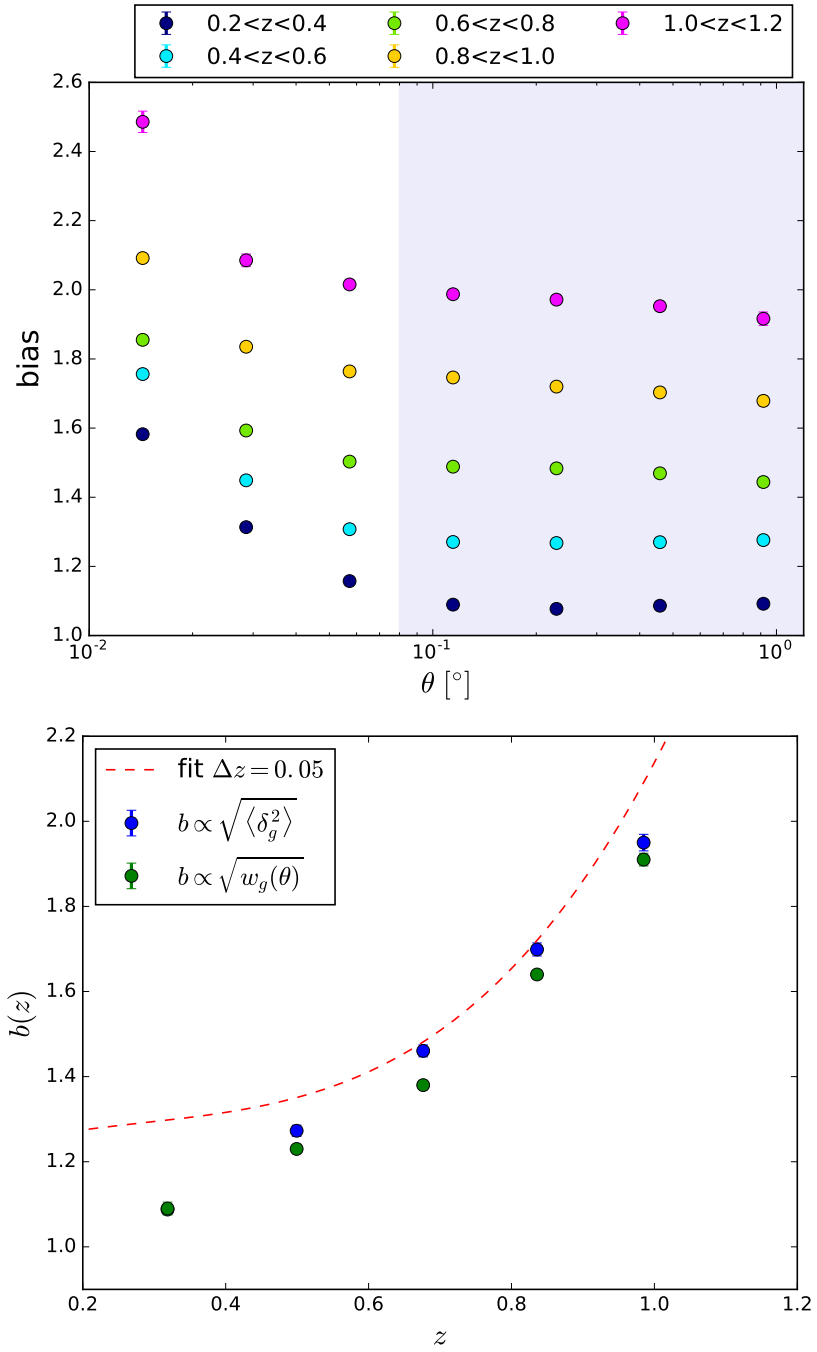


Figure 4.19: Upper panel: Bias from second order CiC as a function of the cell scale obtained in MICE simulation with broad redshift and Gaussian photo-z redshift bins ($\Delta z = 0.2$ and $\sigma_z = 0.05(1+z)$) for different redshift bins. The shaded region corresponds to the points that are in the linear regime $\theta > 0.1^o$. The linear bias is obtained with equation (4.7) with MICE as the data and Buzzard as the simulation. The bias of the Buzzard simulation is obtained fitting theoretical two-point correlation functions $w(\theta)$ with different bias factors and minimizing a χ^2. Lower panel: Bias obtained from second order CiC compared with the one obtained fitting theoretical angular correlation functions with different bias factors. The points from CiC correspond to the average of the points in the shaded region of the upper panel. The dashed line corresponds to the fit of the bias of thin redshift bins ($\Delta z = 0.05$) in the lower panel of Figure 4.12.

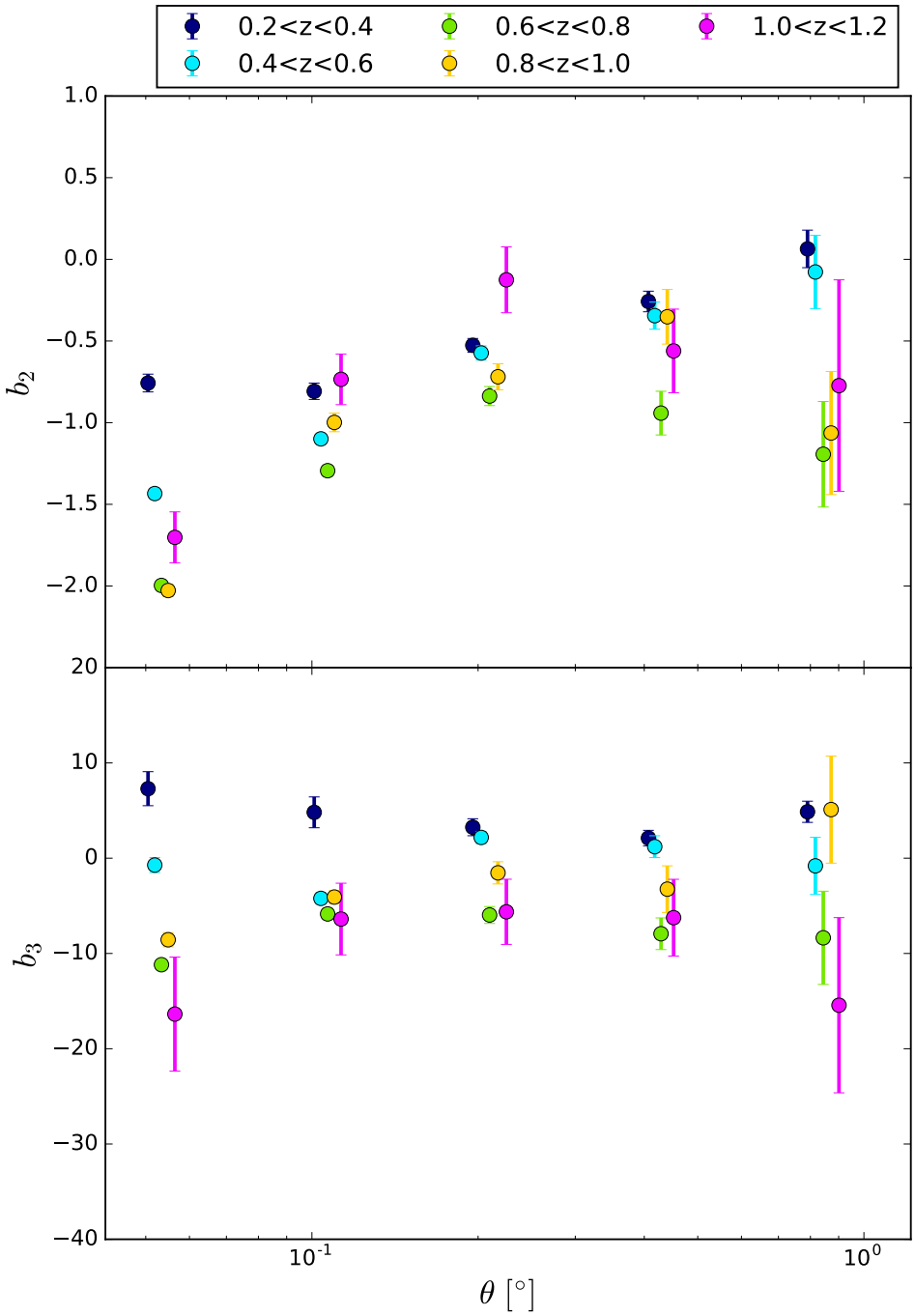


Figure 4.20: Non-linear bias parameters as a function of the cell scale obtained in the Mice simulation with broad redshift and Gaussian photo-z redshift bins ($\Delta z = 0.2$ and $\sigma_z = 0.05(1+z)$) from third and fourth order CiC for different redshift bins.

as the simulation. As the area is small we take the average of 16 different realizations of the Buzzard simulation, to avoid cosmic variance. The points in the lower panel correspond to the average of the points in the shaded region of the upper panel, the linear region where the bias is scale independent. They agree with the bias of the simulation computed minimizing a χ^2 fitting different theoretical correlation functions with different bias. They are of the same order as the fit done to the bias of thin redshift bins ($\Delta z = 0.05$) in the lower panel of Figure 4.12. They do not exactly agree since the photometric redshifts change the redshift distribution and the bias depends strongly on the redshift. Lower redshift bins include the bias of lower redshifts than the bin itself, and then are lower.

We use the linear bias measurements and as explained in section 4.1 we obtain the values of b_2 and b_3 presented in Figure 4.20 using the skewness and kurtosis of the MICE dark matter simulation in the same redshift bins and smeared also with a Gaussian photometric redshift. They are consistent with constant values in the linear scales but the errors are bigger than in previous cases due to the smaller area considered here and the photometric redshift.

4.2.3 Conclusions from MICE simulation

What we have learnt from MICE is that the bias computation method from CiC works in simulations in ideal conditions. As can be seen in Figure 4.13 we can recover the real linear bias with CiC. Taking wider redshift bins we also recover the same bias as in thin redshift bins (Figure A.3). Adding a Gaussian photometric redshift mixes redshifts and increases the errors but the linear bias computation still agrees with the two-point correlation computation (Figure 4.19). When we add the SV mask the area is smaller and we see cosmic variance effects, which result in bigger errors. We were also able to obtain the non-linear bias parameters consistent with constant values in linear scales. Then we can conclude that the bias computation method works even with the SV mask and a photometric redshift.

4.3 Results for DES-SV data

We have tested the methodology with simulations, we have checked that the method works even with a mask and a photo-z. Now we apply our CiC procedure to extract the galaxy bias to DES data. We do CiC with the DES data sample presented in Section 3.8, the so-called Benchmark sample. Using the same footprint, selection cuts and redshift bins as in [142], $z \in [0.2,0.4],[0.4,0.6],[0.6,0.8],[0.8,1.0],[1.0,1.2]$, we do CiC using TPZ photo-z. Figure 4.21 shows the density contrast maps for the redshift bin $0.6 < z < 0.8$ and $nside = 128, 512, 1024$. Figure 4.22 shows the moments of the density contrast distribution. As we saw in the simulation when we apply a mask and a photo-z the variance for the different redshift bins are mixed. The skewness and the kurtosis are constant and the different redshift bins are also mixed. We do not see here the pattern of the skewness and the kurtosis at low scales of the MICE simulation, which must be due to the way non-linearities are add to the simulation. Here we take scales from *nside* 128 to 4096 (from 0.46° to 0.014°) as bigger scales have just a few pixels inside the mask.

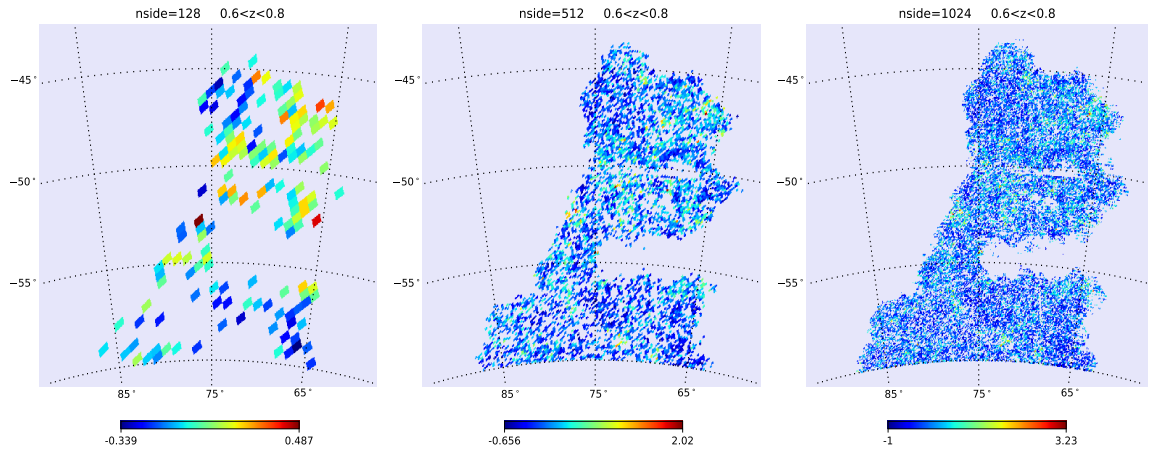


Figure 4.21: Density contrast distribution of the DES SV benchmark sample for the redshift bin $0.6 < z < 0.8$ and $nside = 128, 512, 1024$.

4.3.1 Linear bias

Using the variance, we compute the bias as presented in Section 4.1, with equation (4.7). We compute the ratio of the variance from the SV sample and the moments measured using the MICE simulation emulating the characteristics of the data, i.e. the same color and magnitude cuts, the same mask and the same photo-z (TPZ). In order to emulate the photo-z we smear the true redshift of the galaxies in the simulation using the selection function $\phi(z)$ in Figure 4.23. As the area is small we take the average of 12 different realizations of the MICE simulation to avoid cosmic variance. The bias of the MICE simulation is computed minimizing a χ^2 fitting angular 2pcf with different bias factors to the data. The upper panel of Figure 4.24 shows the bias of the SV sample for different scales and different redshift bins. In the linear regime ($\theta > 0.1^o$) the bias is constant with the scale, except for the lowest redshift bin where non-linearities must be taken into account. The linear scale changes with redshift, being higher for low redshifts due to non-linear gravitational collapse. In [142] they study the dependence of the linear scale with the redshift and find that it's always below $\theta = 0.1^o$. In order to compare with the values found in other DES analyses, we compute the average of the bias results in the linear scales and represent them in the lower panel of Figure 4.24. The errors include the systematic errors that are analysed in detail in Section 4.3.3. We have taken the bias at the mean redshift of the photometric redshift bin weighted with its $\phi(z)$ for the CiC and the 2pcf measurements. In this Figure we see that the values are compatible with the two-point correlation study [142], the CMB-galaxy cross-correlations study [178], and the galaxy-galaxy lensing [179].

4.3.2 Non-linear bias

It is clear that non-linearities must be taken into account, they can not be neglected at low redshifts or small scales. To obtain the values of b_2 and b_3 we use the linear bias measurements from the previous section and as explained in section 4.1, we compute the values of b_2 and b_3 with equation (4.8) with the MICE dark matter simulation values for the skewness

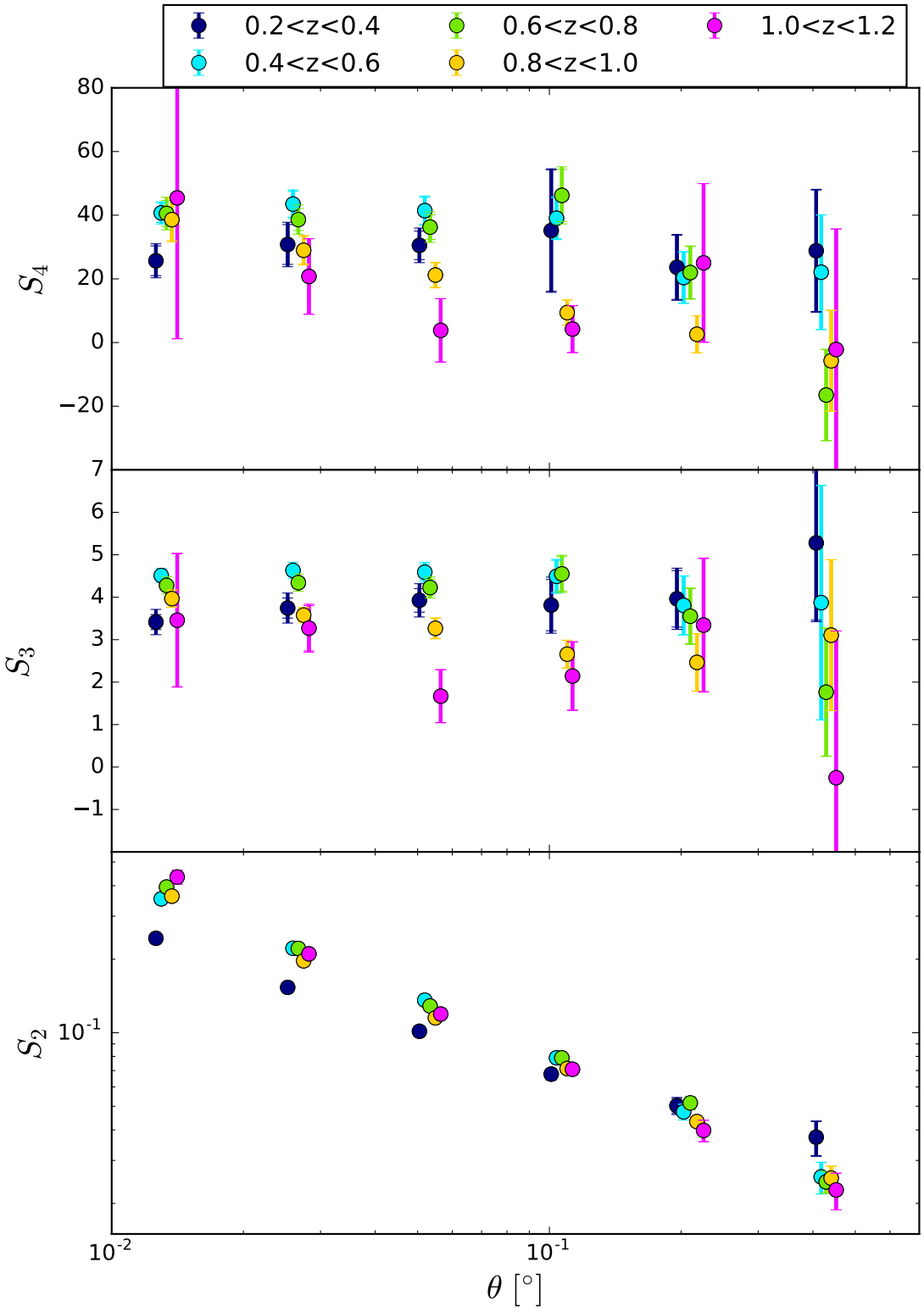


Figure 4.22: Moments of the density contrast distribution of the DES SV sample with TPZ for five different redshift bins and different scales.

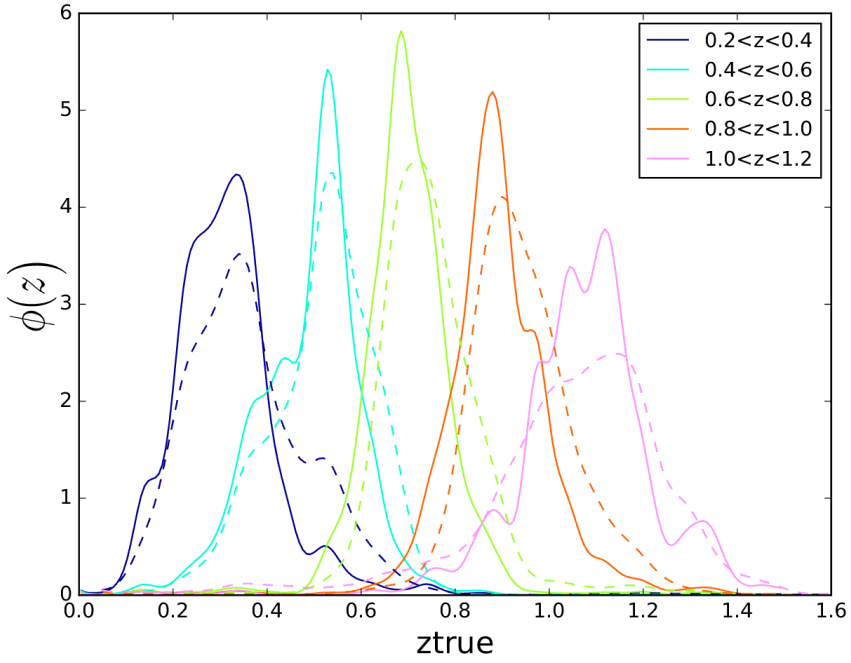


Figure 4.23: Redshift distribution of the galaxies in each photometric redshift bin using TPZ (solid lines) and BPZ (dashed line) in DES-SV benchmark data.

and the kurtosis of dark matter. We can use the Skewness and the Kurtosis of the MICE dark matter simulation, as they hardly depend on cosmology (see equation 2.81). The values are shown in Figure 4.25 and are compatible with constant values within the errors. For the last redshift bin the errors are big and we are not sensible to b_2 and b_3. In Figure 4.26 we compare the $c_2 = b_2/b$ values with the ones obtained in DR5 of SDSS by [123], which are between the dotted lines. They are of the same order, we do not expect them to be exactly the same as the linear and non-linear bias depend strongly on the galaxy population.

4.3.3 Systematics

In this section, we explore the effects that several potential sources of systematic uncertainty have on our moment measurements. Since our main observable is related to the number of galaxy-counts in a given redshift interval, we are interested in observational effects that can affect this number. The main potential sources of systematic uncertainty are changes in airmass, seeing, sky brightness, star-galaxy separation, galactic extinction and the possible errors in the determination of the photometric redshift. In order to evaluate the effect of these, we use the maps introduced at [180], that can be seen in Figure 4.27. To account for the star-galaxy separation we proceed as in [142] and use the USNO-B1 catalogue [181]. We also use the SFD dust maps introduced at [182]. Figures A.12, A.13 and A.14 show the moments of the density contrast distribution for different scales and redshift bin $0.6 < z < 0.8$ taking out the 5% and 10% of the pixels with higher airmass, fwhm and skybrightness respectively. We plot the relative error under the moments. We see that except for same cases explained later, the differences are inside the statistical errors. The bigger differences are for the biggest scales where the statistics are small and

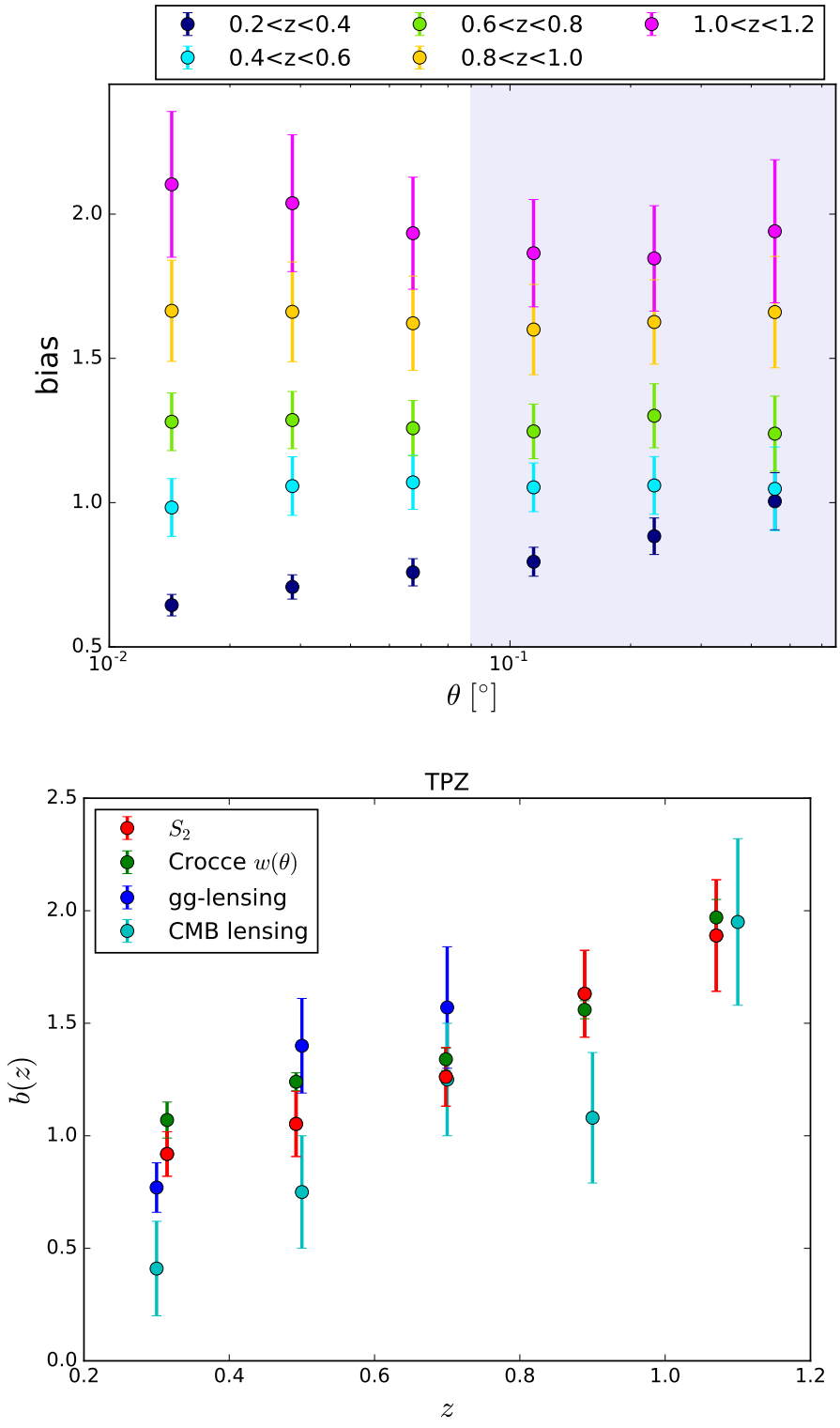


Figure 4.24: Upper panel: Bias obtained for the SV benchmark sample with TPZ from second order CiC as a function of the cell scale. Linear bias is obtained with equation (4.7). The bias of the MICE simulation is obtained fitting theoretical two-point correlation functions $w(\theta)$ with different bias factors and minimizing a χ^2. Lower panel: Bias obtained from second order CiC compared with the two-point correlation study [142], the CMB-galaxy cross-correlations study [178], and the galaxy-galaxy lensing [179]. The points correspond to the average of the points in the shaded region of the upper panel, we consider the linear region above $\theta = 0.1^o$.

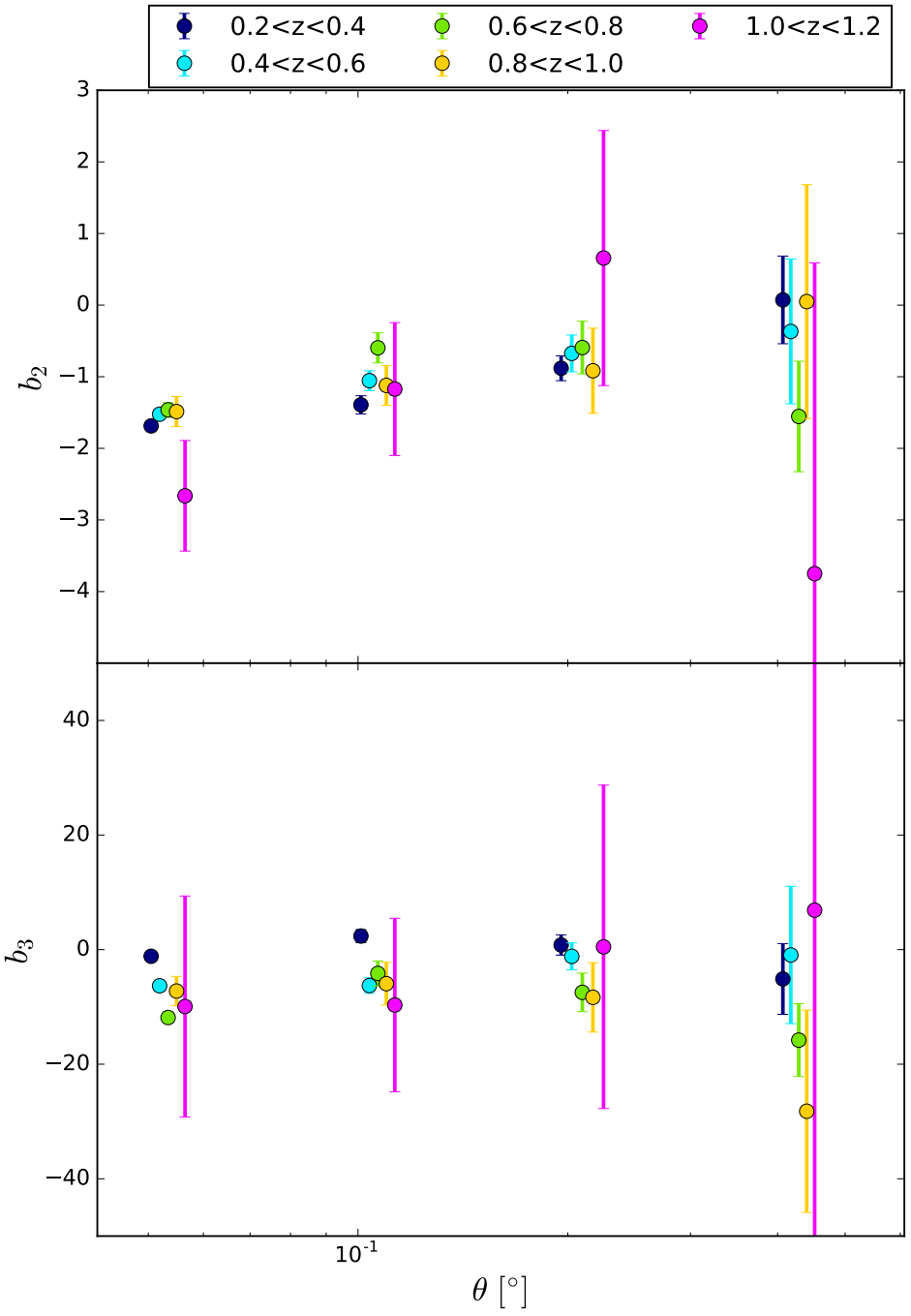


Figure 4.25: Non-linear bias parameters as a function of the cell scale for different redshift bins obtained in the DES-SV sample from third and fourth order CiC with equations (4.8) with the MICE dark matter simulation values for the dark matter skewness and kurtosis.

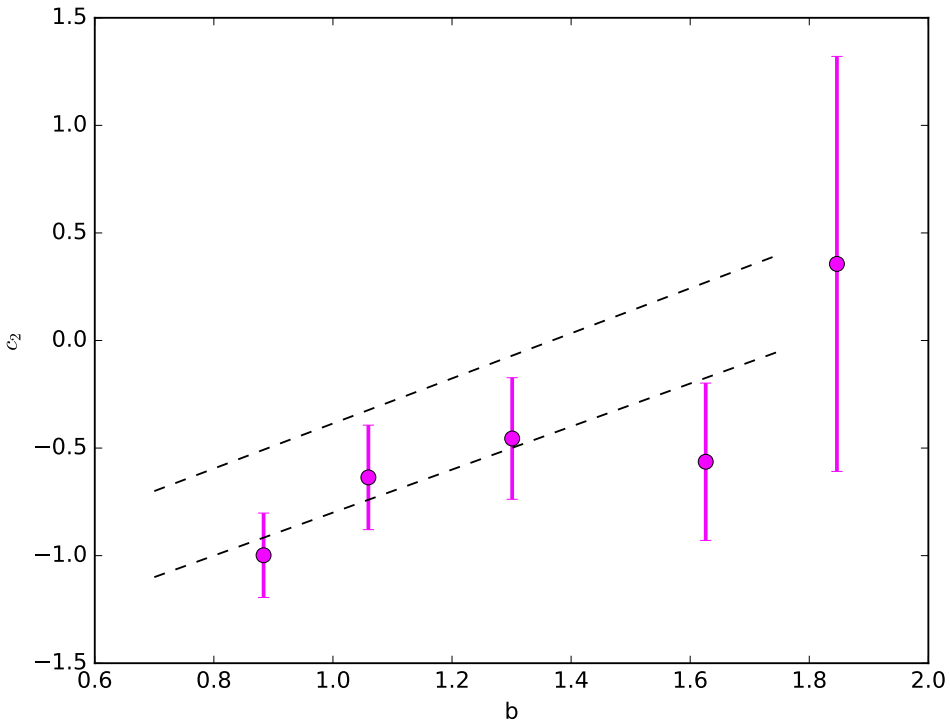


Figure 4.26: Non-linear bias $c_2 = b_2/b$ as a function of the linear bias for $nside = 256$. The dotted lines are from [123].

taking out some pixels makes a difference.

What follows is a detailed step-by-step guide to our analysis: we select one of the aforementioned maps and locate the pixels where the value of the systematic is below the percentile level t. We compute the moments of the density contrast distribution in these pixels, and their respective errors using bootstrap. We change the threshold to $t+5$, repeat the process, and evaluate the difference between the moments divided by the moments in the original footprint $\Delta S_i(t)/\langle S_i \rangle$. An example of the results of this procedure can be found in Figure A.11. We consider that a systematic effect is present if the average of $\Delta S_i(t)/\langle S_i \rangle$ is different than zero at 2σ confidence level or above for the different values of t from the 50th tile to the 100th tile. Then, we assign a systematic uncertainty equal to the value of this average. To be conservative, we consider these effects as independent, so we add these effects in quadrature. We summarize the main systematic effects observed in each redshift bin of our sample:

- Bin $0.2 < z < 0.4$:
  - Seeing in i-band: we assign a 3% systematic uncertainty in S_4.
  - Seeing in z-band: we assign a 2.5% systematic uncertainty in S_4.
  - Sky-brightness r-band: we assign a 1% systematic uncertainty in S_4.
  - Sky-brightness i-band: we assign a 1% systematic uncertainty in S_4.
  - Airmass in g-band: we assign a 1% uncertainty in S_4.
  - Airmass in r-band: we assign a 1% uncertainty in S_4.
  - Airmass in i-band: we assign a 1% uncertainty in S_4.

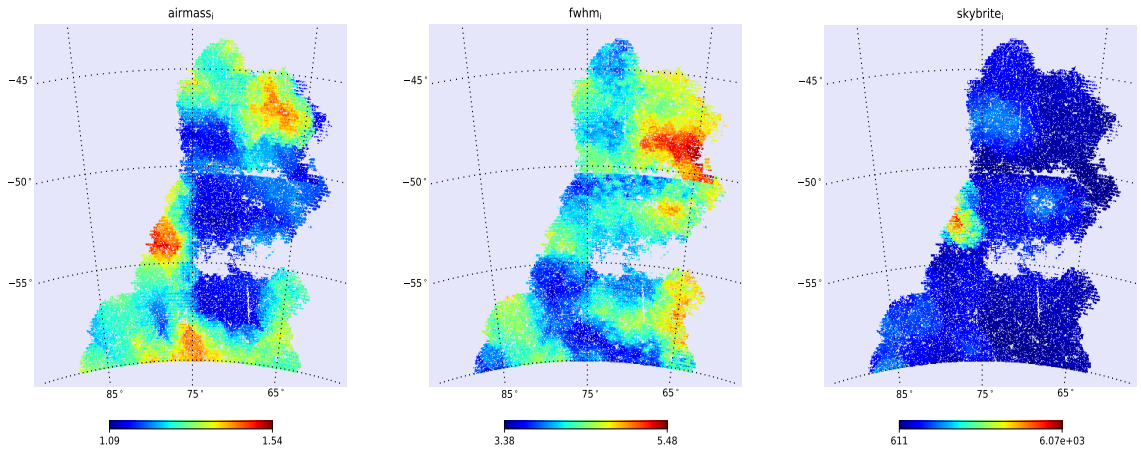


Figure 4.27: Systematic maps of airmass, seeing and skybrightness for the i band.

- USNO-B stars: We assign a 4% uncertainty to S_2, 7% uncertainty to S_3, and 9% to S_4.

- Bin $0.4 < z < 0.6$:
  - Seeing in z-band: We assign a 1.5% uncertainty to S_4.
  - USNO-B stars: We assign a 4% uncertainty to S_2, 3% uncertainty to S_3, and 4% to S_4.
- Bin $0.6 < z < 0.8$:
  - Seeing in g-band: We assign a 2% to S_4.
  - Seeing in r-band: We assign a 2% to S_4.
  - Sky-brightness i-band: We assign a 1.5% uncertainty to S_3, and 3% systematic uncertainty to S_4.
  - Airmass in g-band: We assign a 2.5% uncertainty to S_4.
  - Airmass in r-band: We assign a 2% uncertainty to S_4.
  - Airmass in z-band: We assign a 1.5% uncertainty to S_3, and 3% uncertainty to S_4.
  - USNO-B stars: We assign a 3% uncertainty to S_3, and 5% uncertainty to S_4.
- Bin $0.8 < z < 1.0$:
  - Seeing in g-band: We assign a 2% uncertainty to S_4.
  - Sky-brightness in i-band: We assign a 2% uncertainty to S_3, and a 3.5% uncertainty to S_4.
  - Airmass in g-band: We assign a 2% uncertainty to S_4.
  - Airmass in r-band: We assign a 3% uncertainty to S_4.
  - USNO-B stars: We assign a 3% uncertainty to S_4.
- Bin $1.0 < z < 1.2$:
  - The measurement of S_4 in this bin is dominated by systematics.
  - Sky-brightness i-band: We assign 2% to S_3.
  - Sky-brightness z-band: We assign 3% to S_3.
  - USNO-B stars: We assign a 4.5% uncertainty to S_3.

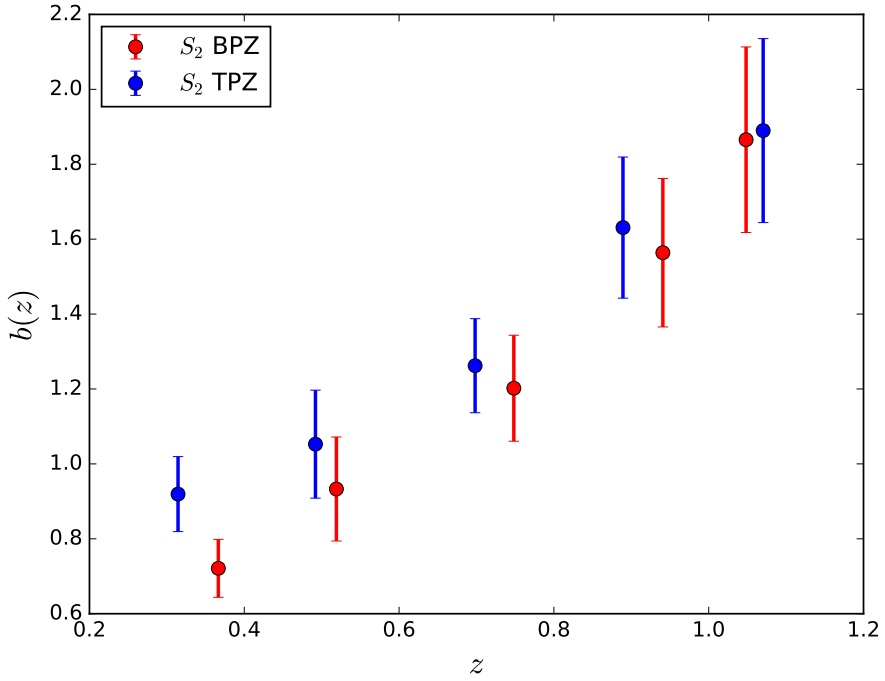


Figure 4.28: Linear bias obtained from second order CiC for SV for two different photometric redshifts, TPZ (blue) and BPZ (red).

Photometric redshift

Photometric redshift is one of the main systematics in photometric surveys like DES. We have repeated the analysis in SV for another photometric redshift, BPZ [165]. In Figure 4.28 we see the values of the bias obtained from CiC for TPZ and BPZ. We have taken the bias at the mean redshift of the photometric redshift bin weighted with its selection function $\phi(z)$. Both computations agree within the errors.

Fiducial Cosmology

When we compute the linear bias we assume a fiducial cosmology to compute the growth factor of the data (equation 4.7). We have checked for MICE simulation, for which we know the cosmology, that changing the cosmological parameters do not change the bias results. We changed all cosmological parameters and the only one that has an effect on the bias is Ω_m, and the error is below the statistical error. In Figure 4.29 the shaded area represents the different bias values obtained changing $\Omega_m = 0.25$ a 4%. We see that this area is smaller than the statistical errors. We have to take into account also that this parameter is known nowadays with high precision, $\Omega_m = 0.3089 \pm 0.0062$ [3] at 68% confidence level for the base ΛCDM model from Planck CMB power spectra, in combination with lensing and external data (BAO+JLA+H_0). Then the error that we make assuming a fiducial cosmology is negligible.

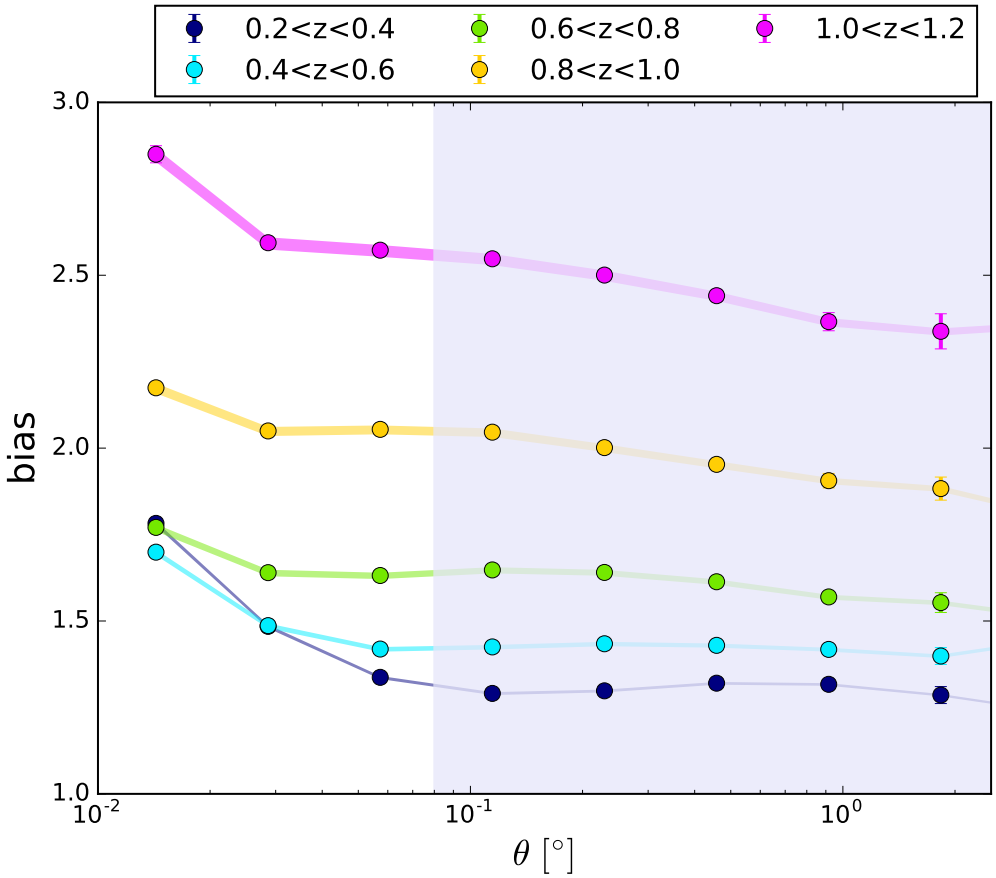


Figure 4.29: Linear bias measurements as a function of the cell scale in MICE simulation from CiC for different Ω_m values in the fiducial cosmology assumed to compute the growth factor of the data in equation (4.7). The shaded area represents the different bias values obtained changing $\Omega_m = 0.25$ a 4%. This area is smaller than the statistical errors.

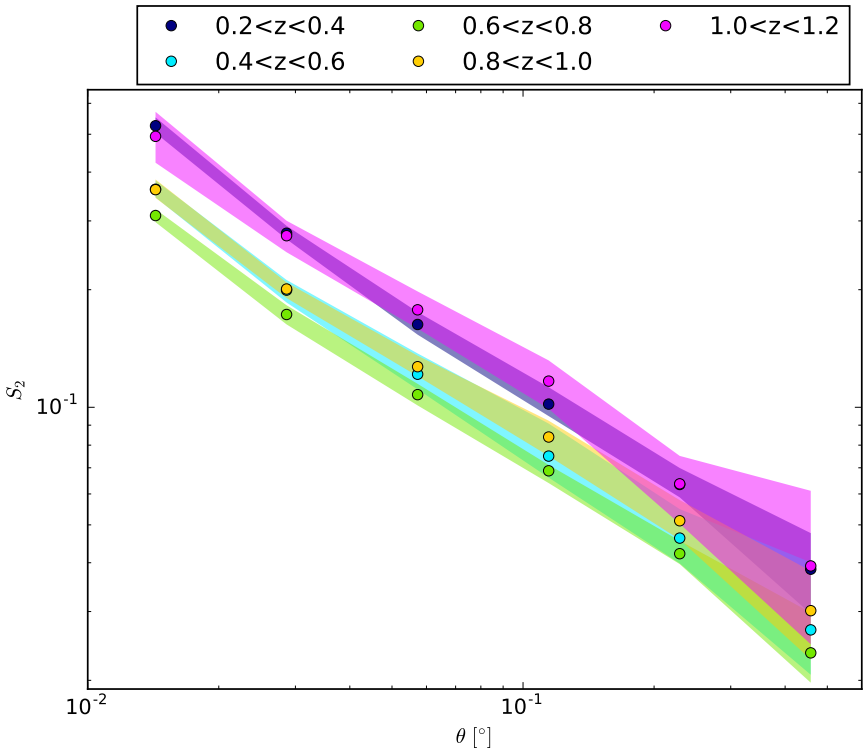


Figure 4.30: Cosmic Variance due to the different values of the variance of the density contrast distribution when we take different realizations of the SV footprint in the MICE octant.

Cosmic Variance

When the area is small we see the effects of cosmic variance. Especially with MICE simulation, since the density is lower than the density in the DES-SV data. The shaded area in Figure 4.30 shows the different values of the variance when we take different realizations of the SV footprint in the MICE octant. We have taken this error into account in the errors computed when using the SV mask.

4.3.4 Lognormal

We use the SV data (Section 3.8) from DES to check the lognormal behaviour of the density contrast distribution. We see that the distributions can be fitted to lognormal distributions with parameters a, σ:

$$P(\delta+1) = \frac{a}{\delta+1} e^{-\frac{(Log(\delta+1)+\frac{\sigma^2}{2})^2}{2\sigma^2}} \tag{4.10}$$

where we have forced $\langle\delta\rangle = 0$, then $\mu = -\sigma^2/2$. The results are shown in Fig. 4.31 for the redshift bin $0.4 < z < 0.6$ and for $nside = 256$ in the upper panel and $nside = 512$ in the lower panel. The authors in reference [183] also find that the galaxy density distribution and the weak lensing convergence (κ_{WL}) are well described by the lognormal distribution using the SV data from DES. Figure 4.32 shows how CiC moments agree with the ones obtained from the lognormal fit at large scales. We added the shot-noise correction to

the lognormal moments computation. For small scales ($nside = 2048,\ 4096$) the density contrast distributions are very discretized (as pixels are very small and we can find up to ~ 10 galaxies per pixel), consequently it is more complicated for the lognormal to fit the data and the moments do not agree.

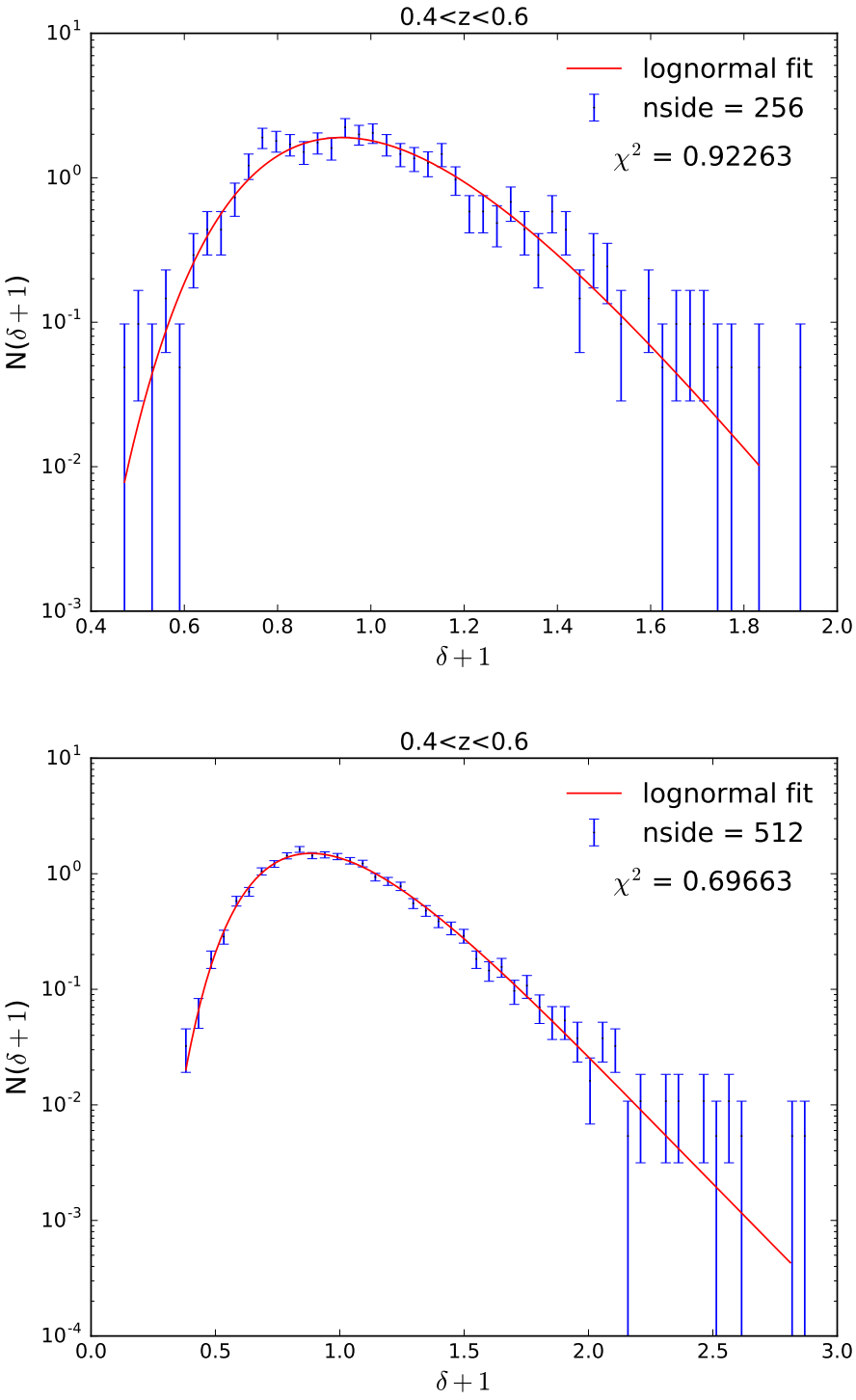


Figure 4.31: Lognormal fit to the density contrast distributions of the DES-SV sample for the redshift bin $0.2 < z < 0.4$ and $nside = 256$ in the upper panel and $nside = 512$ in the lower panel.

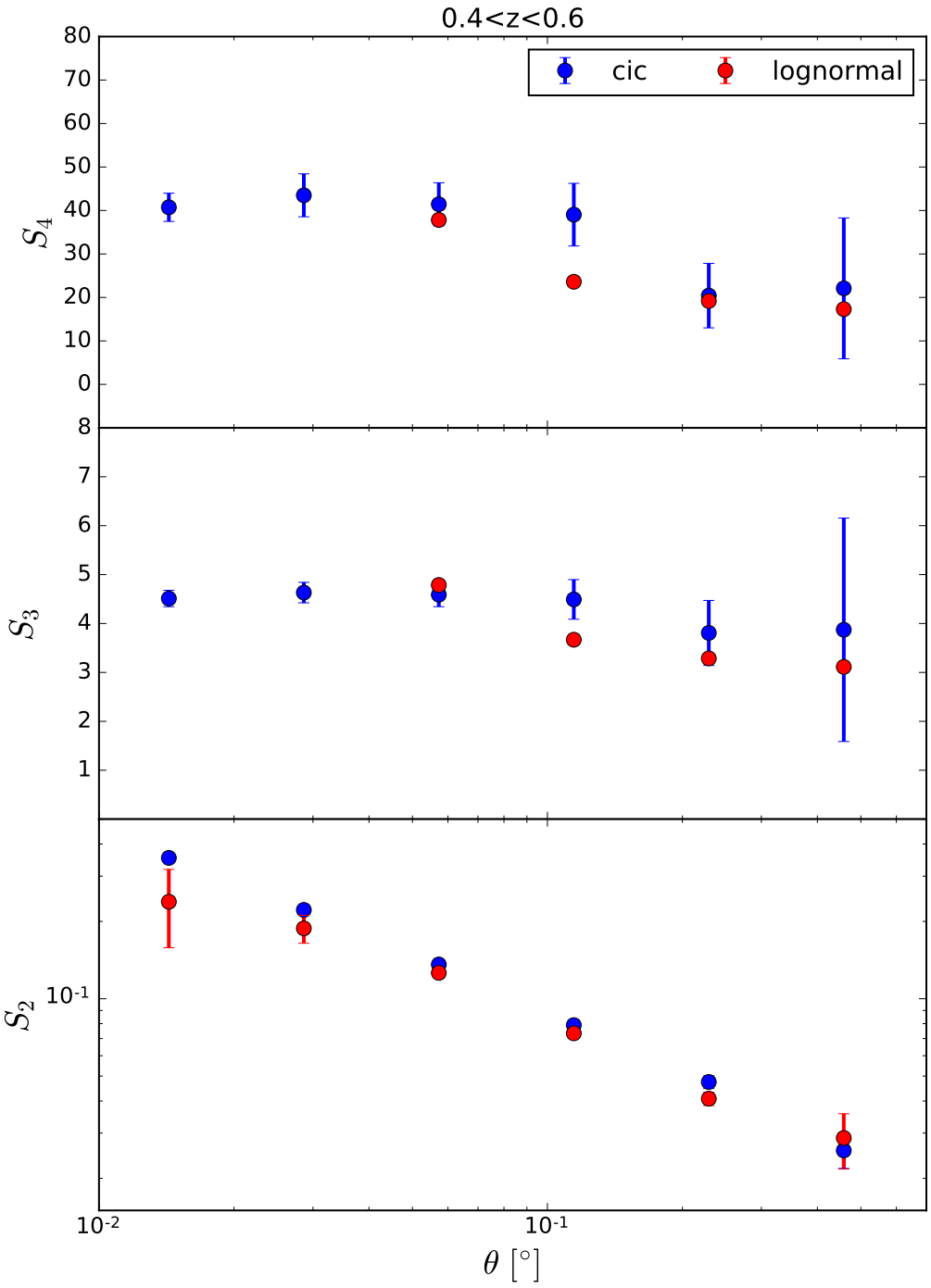


Figure 4.32: Moments as a function of the cell scale of the DES-SV sample computed with CiC (blue) and the lognormal fit (red), for the redshift bin $0.2 < z < 0.4$.

5. Conclusions

In recent years, photometric redshift galaxy surveys have arisen as powerful probes of the Large Scale Structure (LSS) of the Universe and dark energy. The Dark Energy Survey (DES) is one of these surveys. It is a Stage III survey and is now finishing their observations and will publish their final results in a few years. Its biggest drawback is the lack of precision in the direction of the line-of-sight and the systematic effects that are associated with it. Well constrained systematic effects and robust observables are required in order to exploit the best performance of such surveys. Simple observables like the galaxy Counts-in-Cells (CiC) provide information about the LSS and gives an estimate of how different systematic effects can affect the measurements.

A good measurement of the galaxy bias is essential for redshift surveys, as it can introduce a systematic effect on the determination of cosmological parameters. The galaxy bias is highly-degenerate with other cosmological parameters. An independent method to determine the galaxy bias allows to break these degeneracies and improve the overall sensitivity to the underlying cosmology. In this thesis we have developed a method to extract the galaxy bias from Counts-in-Cells (CiC).

CiC is a method based on dividing the sphere in cells of the same volume and counting the number of galaxies in each cell. We can compute the density contrast in each pixel and then the moments of the distribution. As DES is a photometric survey we do angular CiC, due to the lack of precision in the line of sight. We do CiC with healpix pixels, as it is easier to deal with the mask and systematic effects written in this scheme. We verified that, even though the moments of the density contrast distribution depend on the shape of the cell, this difference is negligible as long as the pixel is a regular polygon and we avoid mixing scales and boundary effects. When we use a mask we make sure that we do not mix scales taking only pixels with a certain area (90%) inside the mask.

We compute the linear bias from the variance of the density contrast distribution and the higher-order bias parameters (b_2 and b_3) from the third- and fourth- order moments. If we do not have access to a dark matter simulation we cannot address the dark matter variance. Then we compare the bias of the data to the one in a galaxy simulation with the

same characteristics. At the end of our study we had access to the dark matter simulation used in [132]. In order to compute higher-order bias parameters, we use the higher-order moments of dark matter from the MICE dark matter simulation. The third and fourth order moments are hardly cosmology dependent [89], that is why we can use these values for any sample with different cosmology.

We use MICE and Buzzard simulations to test our method and then perform measurements on the public Science Verification data from the Dark Energy Survey. Galaxy bias is highly environmental dependent, it depends on the galaxy population and on the redshift. Then, we make sure that the data and the simulation we are using to compare with have the same magnitude cuts, the same mask, and the same photometric redshift.

We checked that our linear bias measurement from CiC agrees with the real bias with MICE simulation. Figure 4.13 shows an agreement between our measurement and the one from [132], obtained comparing the angular two-point correlation function of the data with the one from dark matter. We checked that these results are recovered even when we apply a mask and a Gaussian photometric redshift to the simulation. Then, we obtained the linear bias in the SV data in agreement with previous bias measurements from other DES analyses. In Figure 4.24 we see the CiC values are compatible with the two-point correlation study [142], the CMB-galaxy cross-correlations study [178], and the galaxy-galaxy lensing [179].

Finally, we computed the non-linear bias parameters up to third order in the SV sample (Figure 4.25). b_2 and b_3 are consistent with a constant value within the errors for linear scales (above 0.1°). However, given the uncertainties associated to these quantities, it is difficult to draw any conclusions from them. This will be clarified when more data is available, we will have 5000 square degrees at the end of the survey.

The systematic errors are in general lower than the statistical errors, in agreement with the systematic study done by [142].

In conclusion, CiC is a simple but effective method to obtain the linear and non-linear bias. The strength of this method is that it is not demanding computationally and it is based on a simple observable, the galaxy number counts. The only drawback is that it requires a simulation with the same characteristics as the data to compute the linear bias, and a dark matter simulation to compute the non-linear bias parameters. We have checked that for the public available DES data we can recover the lineal bias in agreement with previous measurements and we have measured the non-linear bias. With future DES data releases there will be more observed area and this results will improve as it has been shown with MICE simulation.

6. Conclusión

En los últimos años los cartografiados fotométricos se están utilizando para investigar la estructura a gran escala del Universo y la energía oscura. El 'Dark Energy Survey' es uno de estos cartografiados. DES pertenece a la fase III, está finalizando ahora la toma de datos y en unos pocos años publicará los resultados finales. Su mayor inconveniente es la falta de precisión en la línea de visión y los sistemáticos asociados a ella. Para explotar al máximo este tipo de cartografiados es necesario restringir al máximo los sistemáticos y utilizar observables robustos. Observables sencillos como Counts-in-Cells (CiC) o 'Cuenta-por-Celdas' da información sobre la estructura a gran escala y puede estimar cómo afectan los diferentes sistemáticos a las medidas.

Una buena medida del bias de las galaxias es esencial para los cartografiados cosmológicos, ya que puede introducir errores sistemáticos en la determinación de los parámetros cosmológicos. El bias de las galaxias está altamente degenerado con otros parámetros cosmológicos, por lo tanto un método independiente para calcularlo ayuda a romper esta degeneración y mejorar la sensibilidad a la cosmología. En esta tesis se ha desarrollado un método para extraer el bias de las galaxias con el método CiC.

CiC es un método basado en dividir la muestra en celdas del mismo volumen y contar el número de galaxias en cada celda. Podemos calcular el contraste de densidad en cada celda y de ahí los momentos de la distribución. Como DES es un cartografiado fotométrico hacemos CiC en la dirección angular, debido a la falta de precisión en la dirección radial. Hacemos CiC con pixeles de healpix, ya que así es más sencillo tratar con una máscara y sistemáticos que utilizan este esquema. Se ha comprobado que aunque los momentos dependan de la forma del pixel, esta diferencia es despreciable siempre y cuando el pixel sea un polígono regular y se evite mezclar escalas y los efectos de borde. Cuando se utiliza una máscara hay que asegurarse de no mezclar escalas, para ello cogemos sólo pixeles con un cierto porcentaje (90%) dentro de la máscara.

Calculamos el bias lineal a partir de la varianza de la distribución del contraste de densidad, y los órdenes superiores del bias (b_2 y b_3) con los momentos de orden tres y cuatro. Si no se tiene acceso a una simulación de materia oscura no podemos calcular la

varianza de su contraste de densidad. Entonces para calcular el bias lineal comparamos la varianza de los datos con la de una simulación de galaxias con las mismas características. Al final de nuestro estudio tuvimos acceso a la simulación de materia oscura utilizada en [132]. En el caso de los órdenes superiores utilizamos la simulación de materia oscura MICE para calcular los momentos de orden tres y cuatro de materia oscura, ya que apenas dependen de la cosmología [89], y pueden utilizarse para cualquier muestra con otra cosmología.

Utilizamos las simulaciones MICE y Buzzard para comprobar nuestro método y medir el bias de las galaxias en la muestra del Science Verification de DES. Como el bias de las galaxias depende fuertemente de la población de galaxias y el redshift, es muy importante que tanto los datos como la simulación con la que comparamos tengan los mismos cortes en magnitud, la misma máscara y el mismo redshift fotométrico.

Se ha comprobado con la simulación MICE que el bias lineal que se obtiene mediante CiC está de acuerdo con el bias real. La Figura 4.13 muestra el resultado de CiC frente al de [132] comparando la función de correlación de galaxias y la de materia oscura. También hemos comprobado que se recuperan resultados coherentes aplicando una máscara y un redshift fotométrico a la simulación. Después hemos aplicado el método a los datos Science Verification de DES y vemos en la Figura 4.24 que están de acuerdo con estudios previos como el de la función de correlación a dos puntos [142], el de la cross-correlación CMB-galaxias [178], y el de galaxy-galaxy lensing[179].

Finalmente calculamos los parámetros no lineales hasta tercer orden en la muestra SV (Figura 4.25). b_2 y b_3 son consistentes dentro de los errores con un valor constante en escalas lineales (por encima de 0.1°). Sin embargo, dadas las incertidumbres asociadas a estas cantidades es dfifícil de precisar. Se espera que estas medidas mejoren y los errores sean menores con futuros datos de DES. Al finalizar el quinto año se habrá observado en un área mucho mayor, serán 5000 grados cuadrados.

Los errores sistemáticos son en general menores que los errores estadísticos, de acuerdo con el estudio [142].

En conclusión, CiC es un método simple pero efectivo para calcular el bias de las galaxias lineal y no lineal. El punto fuerte de este método es que no es costoso computacionalmente y está basado en un observable fácil de obtener, el contaje de galaxias. El único inconveniente de éste método, es que se necesita una simulación con las mismas características que los datos para calcular el bias lineal, y una simulación de materia oscura para calcular el bias no lineal. Hemos comprobado que podemos calcular el bias lineal de los datos públicos de DES disponibles por el momento de acuerdo con estudios previos, y también hemos calculado el bias no lineal. Con los futuros datos de DES se espera que los resultados mejoren al tener más área, como se ha comprobado con la simulación MICE.

A. Appendix

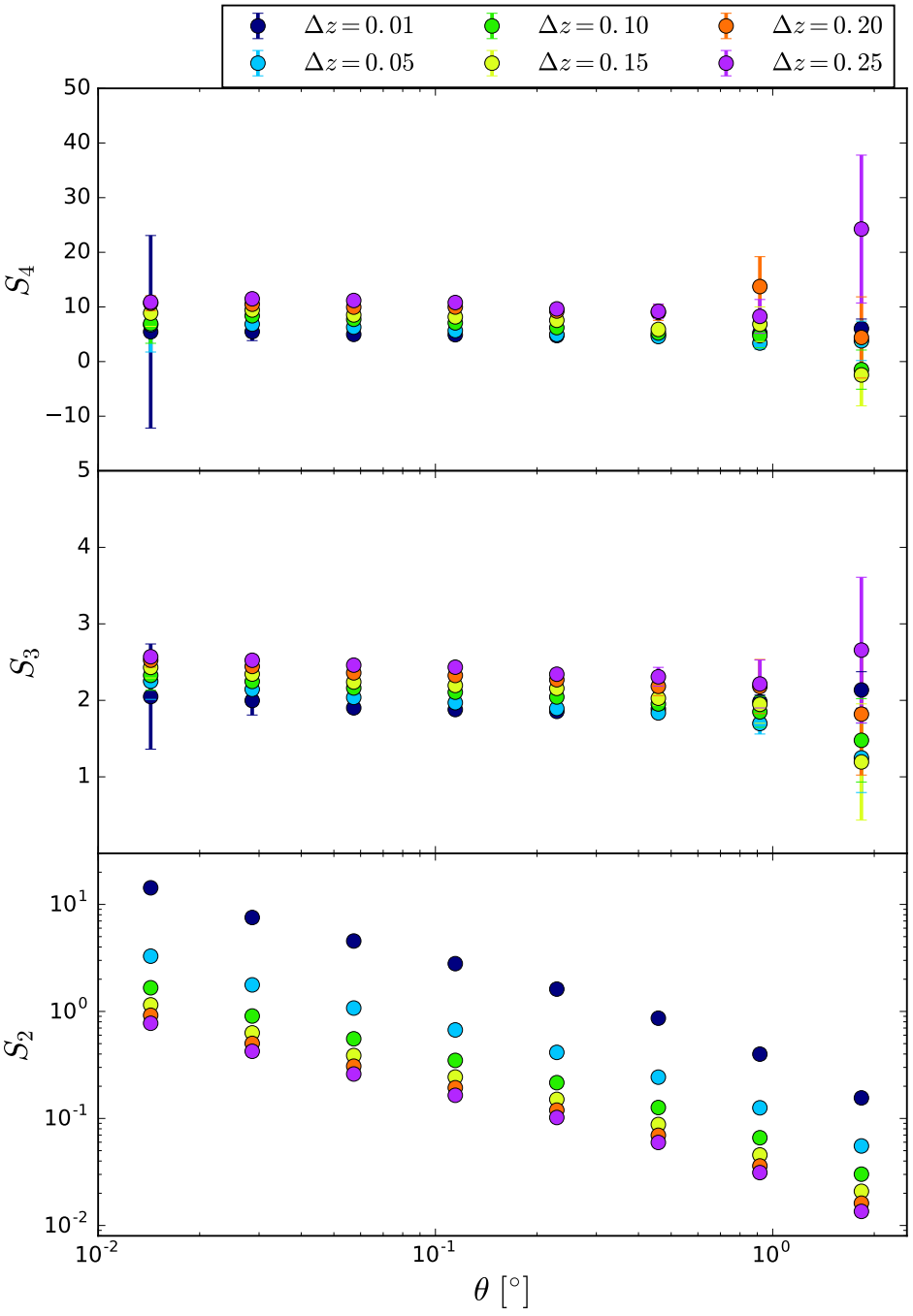


Figure A.1: Moments of the density contrast distribution as a function of the cell scale in the MICE simulation for different redshift bin widths $\Delta z = 0.01, 0.05, 0.10, 0.15, 0.20, 0.25$ centred in $\bar{z} = 1.0$.

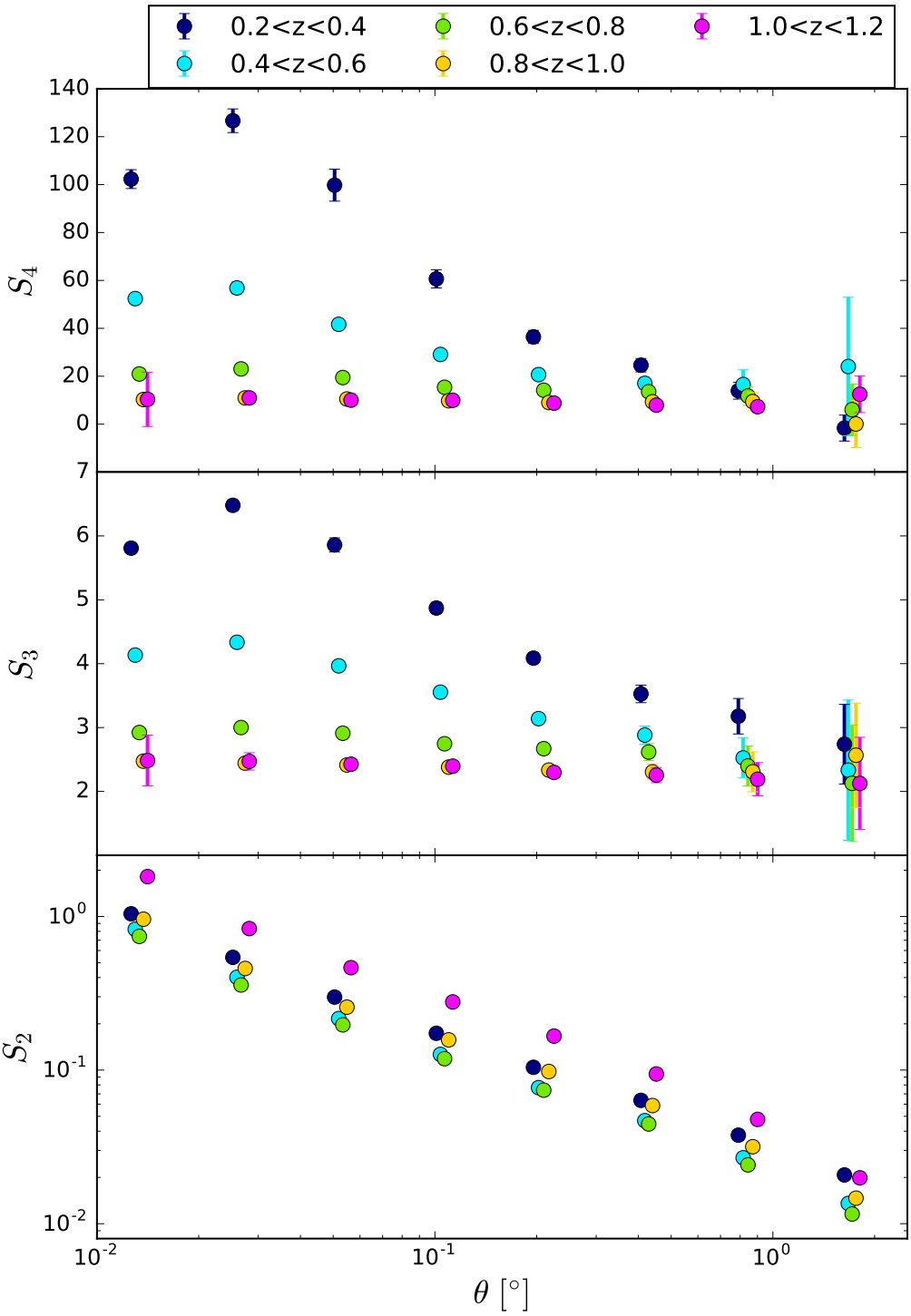


Figure A.2: Moments of the density contrast distribution as a function of the cell scale in the MICE simulation ($\Delta z = 0.2$) for different redshift bins.

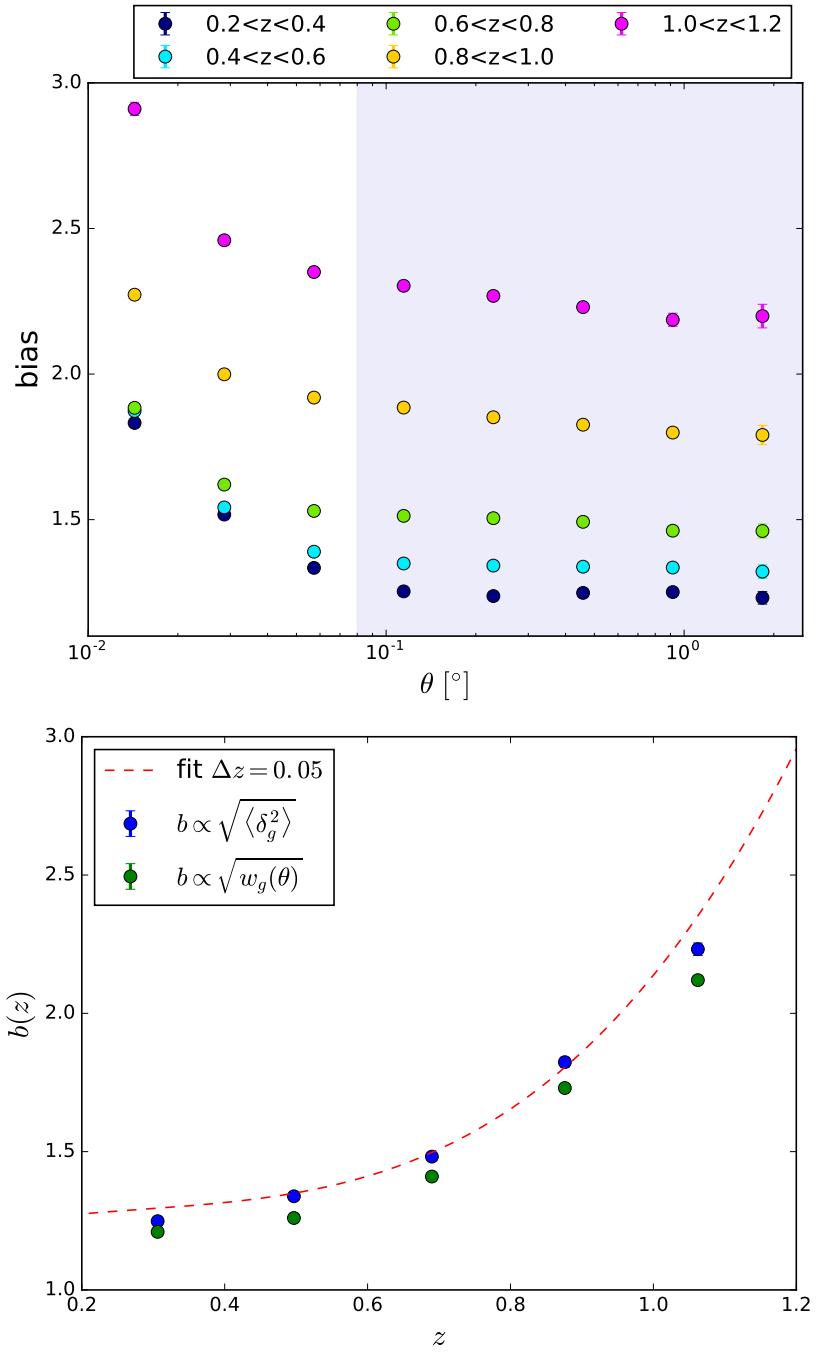


Figure A.3: Upper panel: Bias obtained in MICE simulation ($\Delta z = 0.2$) from second order CiC as a function of the cell scale for different redshift bins. The shadowed region corresponds to the points that are in the linear regime $\theta > 0.1^o$. Linear bias is obtained with equation (4.7) with MICE as the data and Buzzard as the simulation. The bias of the Buzzard simulation is obtained fitting theoretical two-point correlation functions $w(\theta)$ with different bias factors and minimizing a χ^2. Lower panel: Bias obtained from second order CiC compared with the one obtained fitting theoretical angular correlation functions with different bias factors. The points from CiC correspond to the average of the points in the shadowed region of the upper panel. The dashed line corresponds to the fit of the bias of thin redshift bins ($\Delta z = 0.05$) in the lower panel of Figure 4.12.

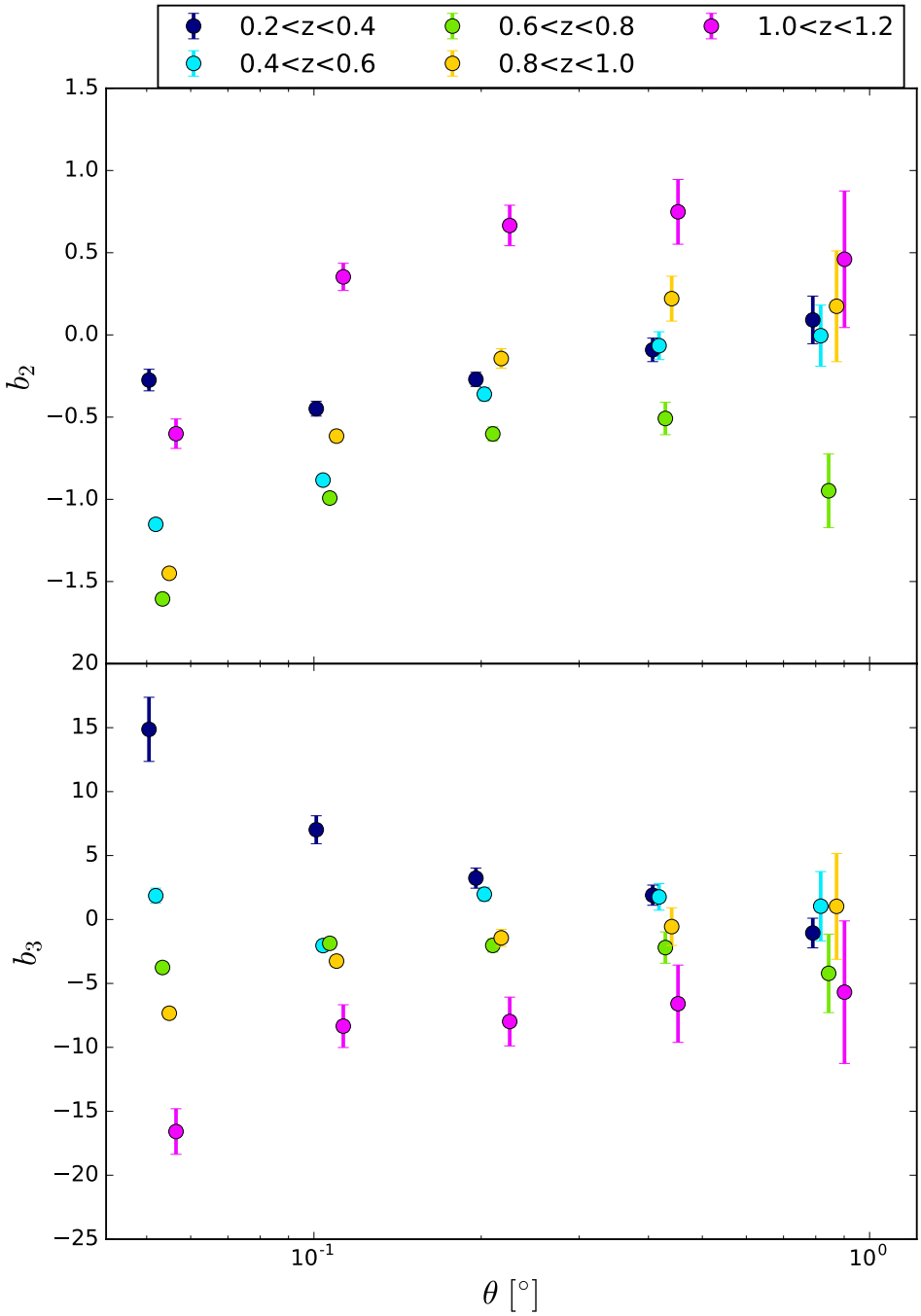


Figure A.4: Non-linear bias parameters as a function of the cell scale obtained in the Mice simulation with broad redshift bins ($\Delta z = 0.2$) from third and fourth order CiC for different scales, and with equation (4.8) and the MICE dark matter simulation values for the dark matter skewness and kurtosis.

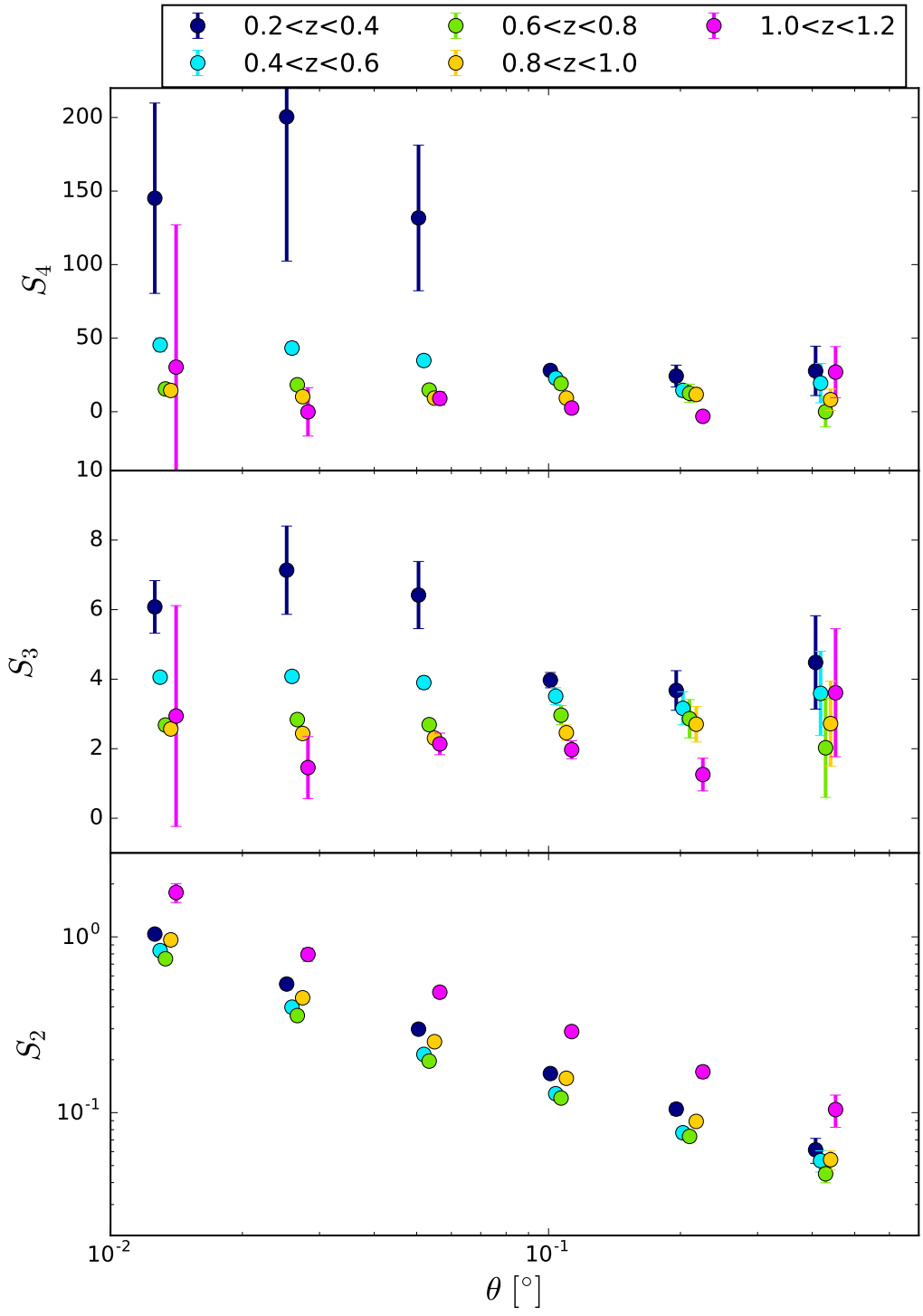


Figure A.5: Moments of the density contrast distribution as a function of the cell scale in the MICE simulation ($\Delta z = 0.2$ and SV mask) for different redshift bins, with equation (4.8) and the MICE dark matter simulation values for the dark matter skewness and kurtosis.

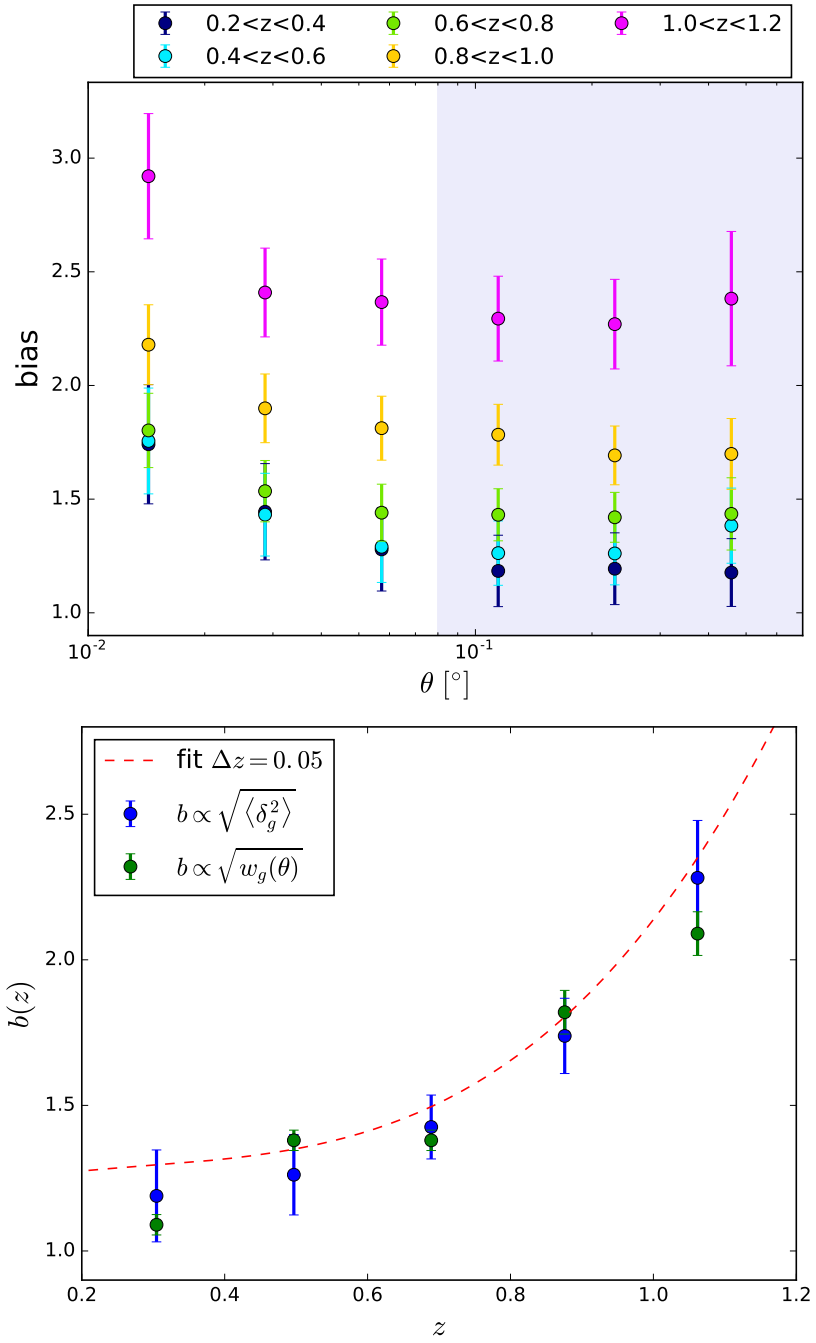


Figure A.6: Upper panel: Bias from second order CiC as a function of the cell scale obtained in MICE simulation with broad redshift bins and SV mask ($\Delta z = 0.2$ and SV mask). The shadowed region corresponds to the points that are in the linear regime $\theta > 0.1^o$. Linear bias is obtained with equation (4.7) with MICE as the data and Buzzard as the simulation. The bias of the Buzzard simulation is obtained fitting theoretical two-point correlation functions $w(\theta)$ with different bias factors and minimizing a χ^2. Lower panel: Bias obtained from second order CiC compared with the one obtained fitting theoretical angular correlation functions with different bias factors. The points from CiC correspond to the average of the points in the shadowed region of the upper panel. The dashed line corresponds to the fit of the bias of thin redshift bins ($\Delta z = 0.05$) in the lower panel of Figure 4.12.

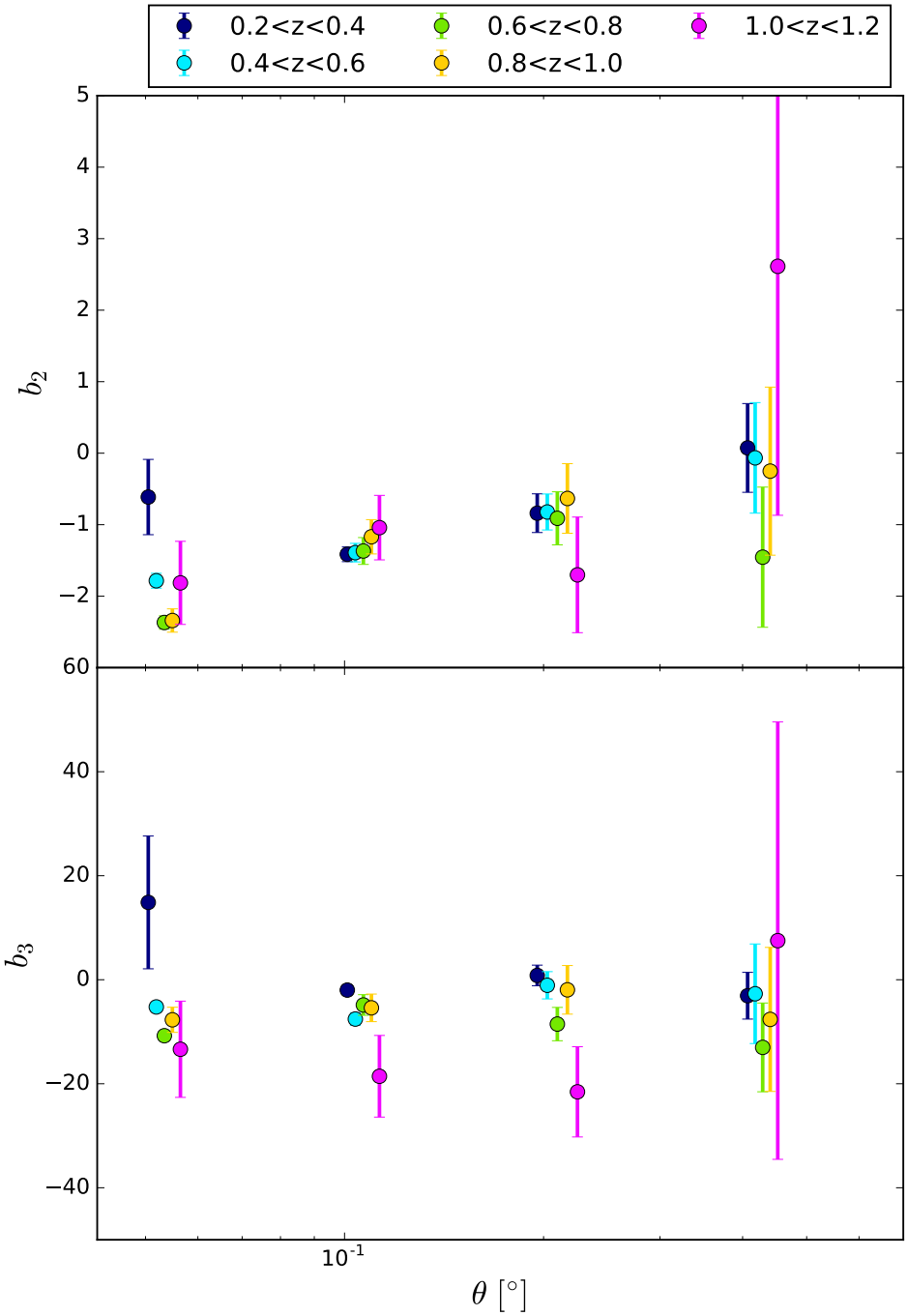


Figure A.7: Non-linear bias parameters as a function of the cell scale obtained in the Mice simulation with broad redshift and SV mask ($\Delta z = 0.2$ and SV mask) from third and fourth order CiC, with equation (4.8) and the MICE dark matter simulation values for the dark matter skewness and kurtosis.

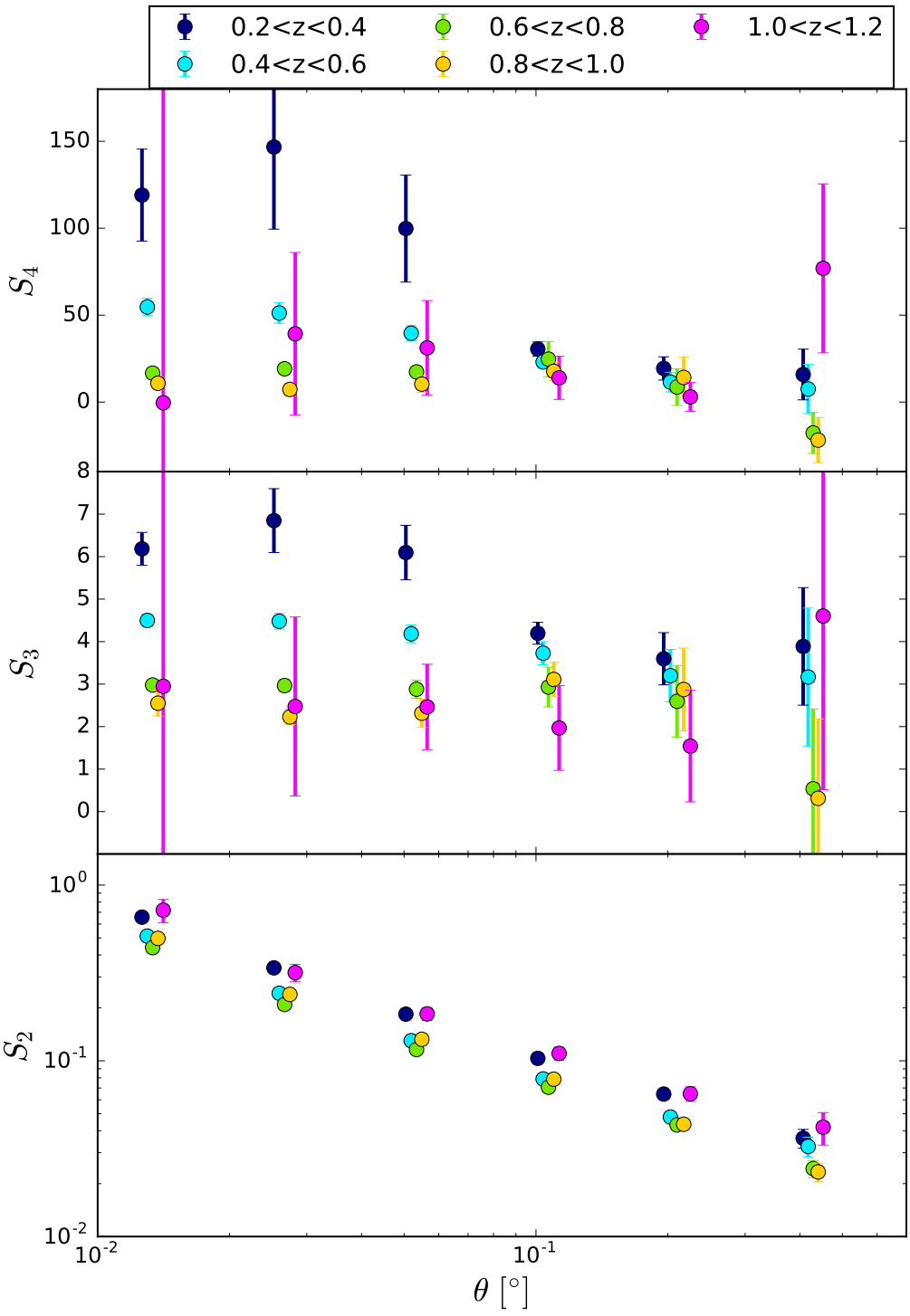


Figure A.8: Moments of the density contrast distribution as a function of the cell scale in the MICE simulation with SV mask and a Gaussian photo-z ($\Delta z = 0.2$ and SV mask and $\sigma_z = 0.05(1+z)$) for different redshift bins.

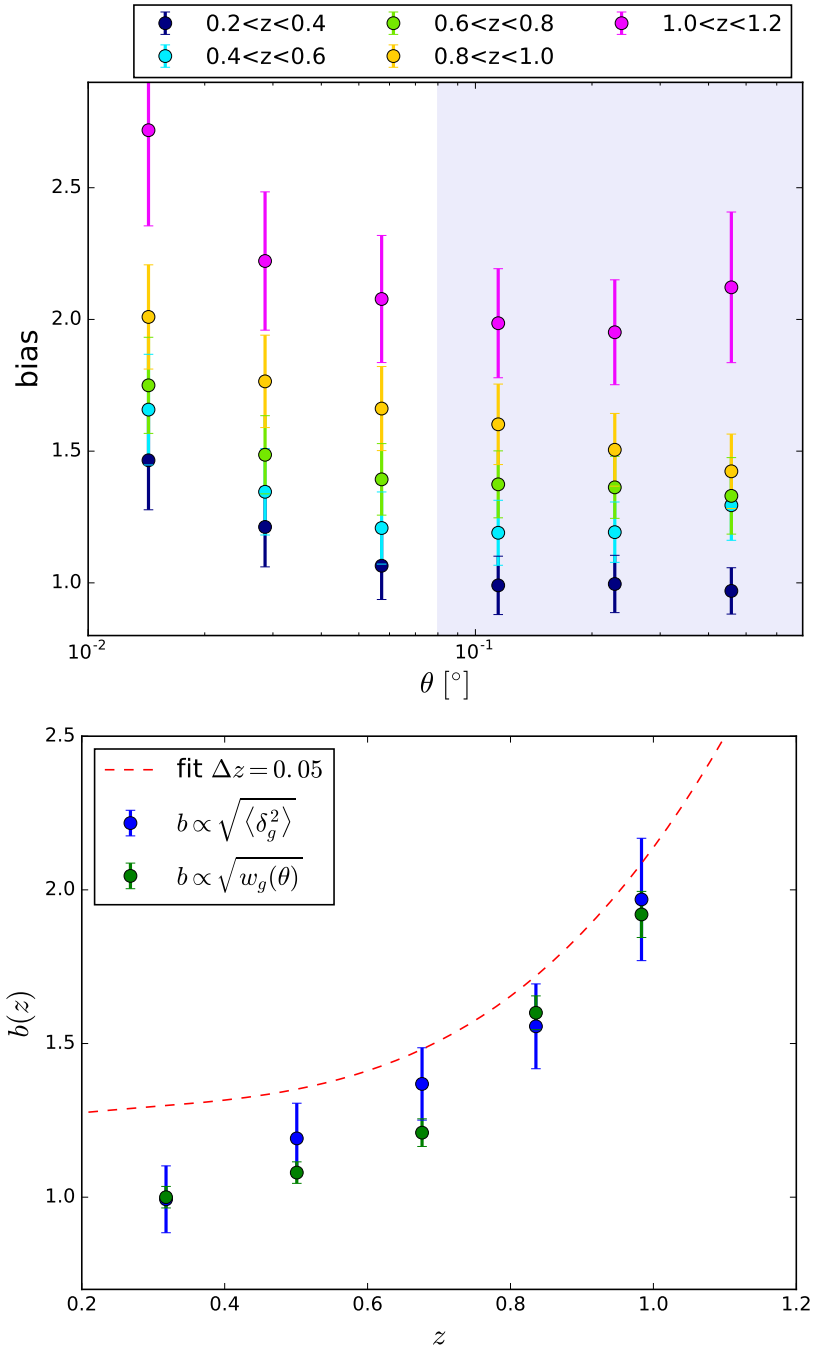


Figure A.9: Upper panel: Bias from second order CiC as a function of the cell scale, obtained in MICE simulation with broad redshift bins, SV mask and a Gaussian photo-z ($\Delta z = 0.2$ and SV mask and $\sigma_z = 0.05(1+z)$). The shadowed region corresponds to the points that are in the linear regime $\theta > 0.1^o$. Linear bias is obtained with equation (4.7) with MICE as the data and Buzzard as the simulation. The bias of the Buzzard simulation is obtained fitting theoretical two-point correlation functions $w(\theta)$ with different bias factors and minimizing a χ^2. Lower panel: Bias obtained from second order CiC compared with the one obtained fitting theoretical angular correlation functions with different bias factors. The points from CiC correspond to the average of the points in the shadowed region of the upper panel. The dashed line corresponds to the fit of the bias of thin redshift bins ($\Delta z = 0.05$) in the lower panel of Figure 4.12.

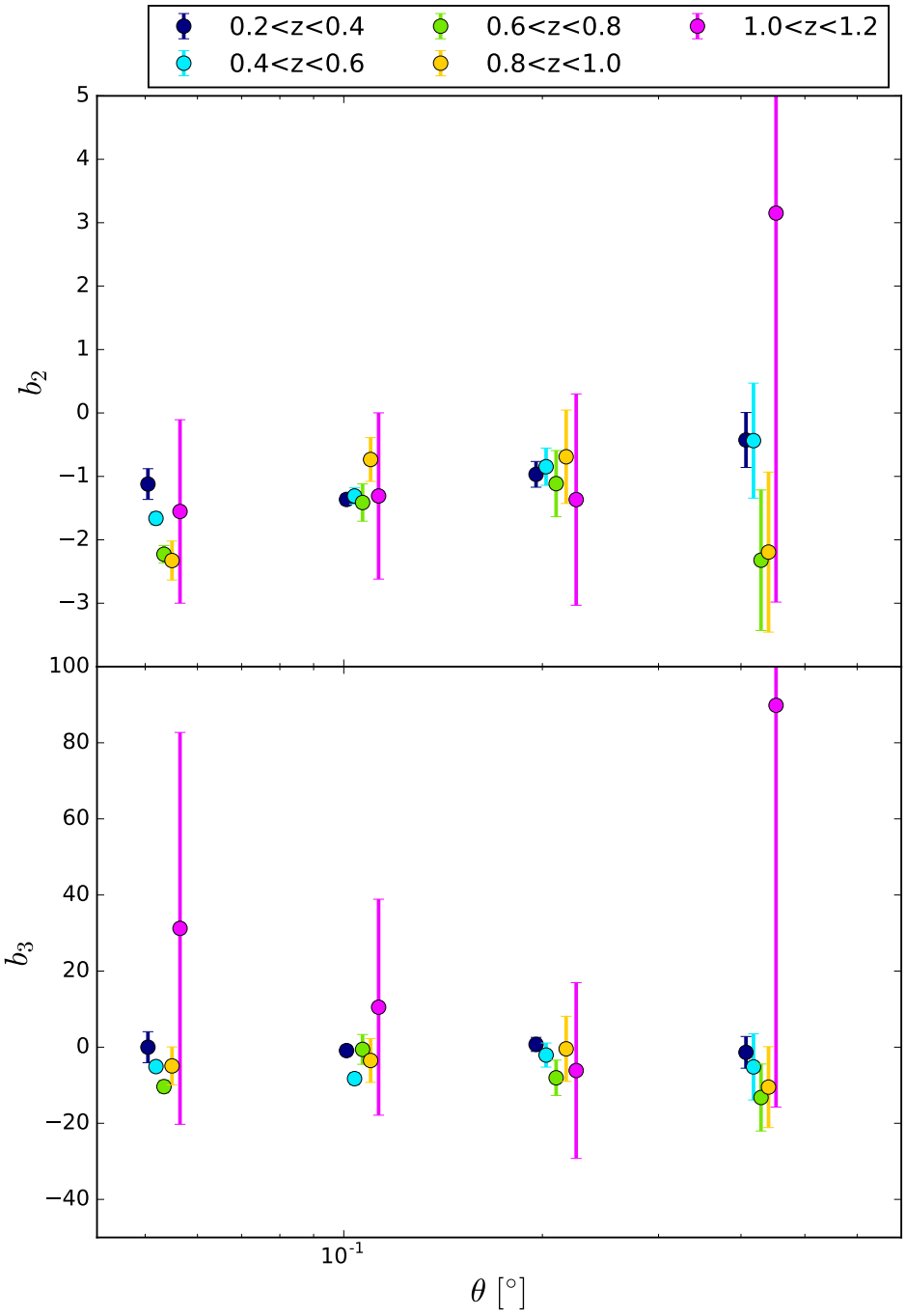


Figure A.10: Non-linear bias parameters as a function of the cell scale obtained in the Mice simulation with broad redshift, SV mask and a Gaussian photo-z ($\Delta z = 0.2$ and SV mask and $\sigma_z = 0.05(1+z)$) from third and fourth order CiC.

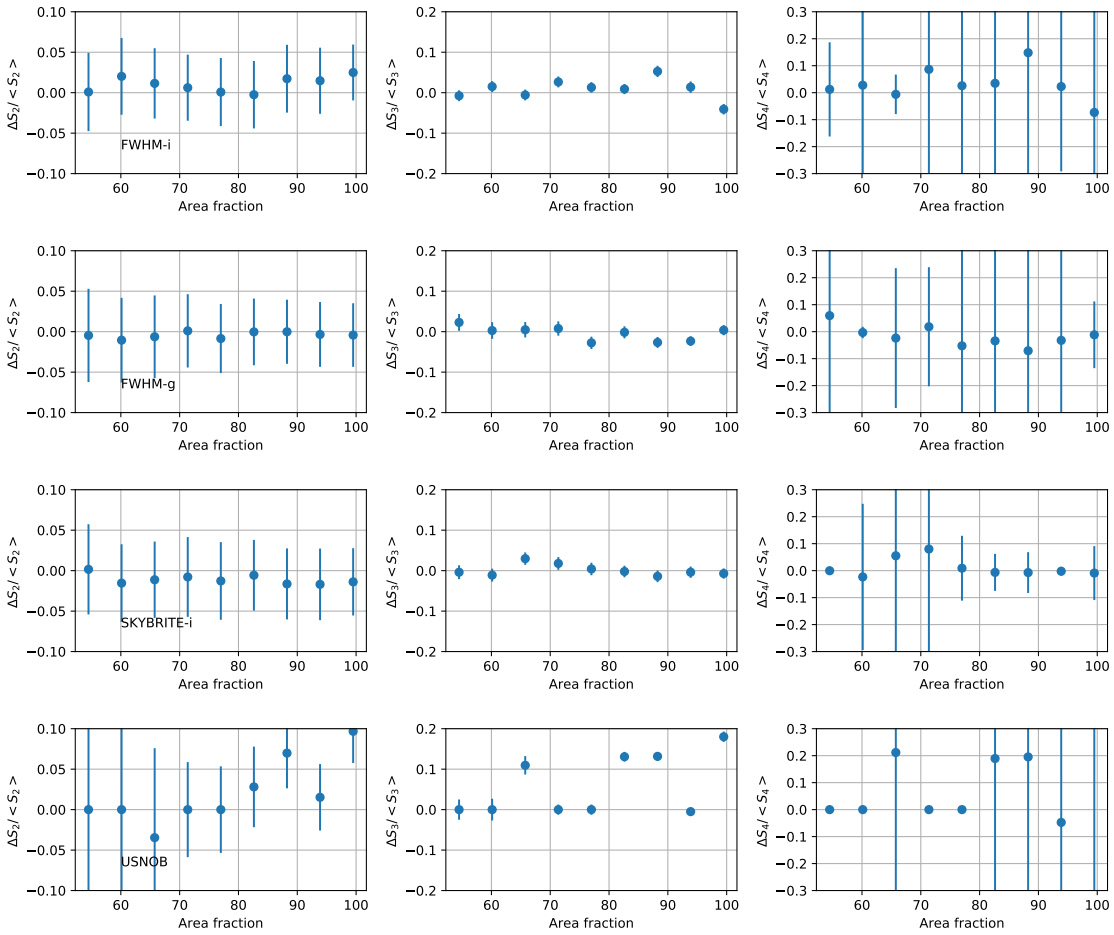


Figure A.11: Dependence of the moments S_i with the variation in the value of potential systematic effects. We show an example for $nside = 2048$ in the redshift bin $0.2 < z < 0.4$ for TPZ. The left column shows the behaviour for S_2, the middle column shows S_3, and the last column shows the results for S_4. The first row corresponds to the results for the seeing in i-band, the second row shows the results for seeing in g-band, the third shows the sky-brightness in i-band. Finally the last row shows the evolution of the moments with the variation in the number of stars per pixel.

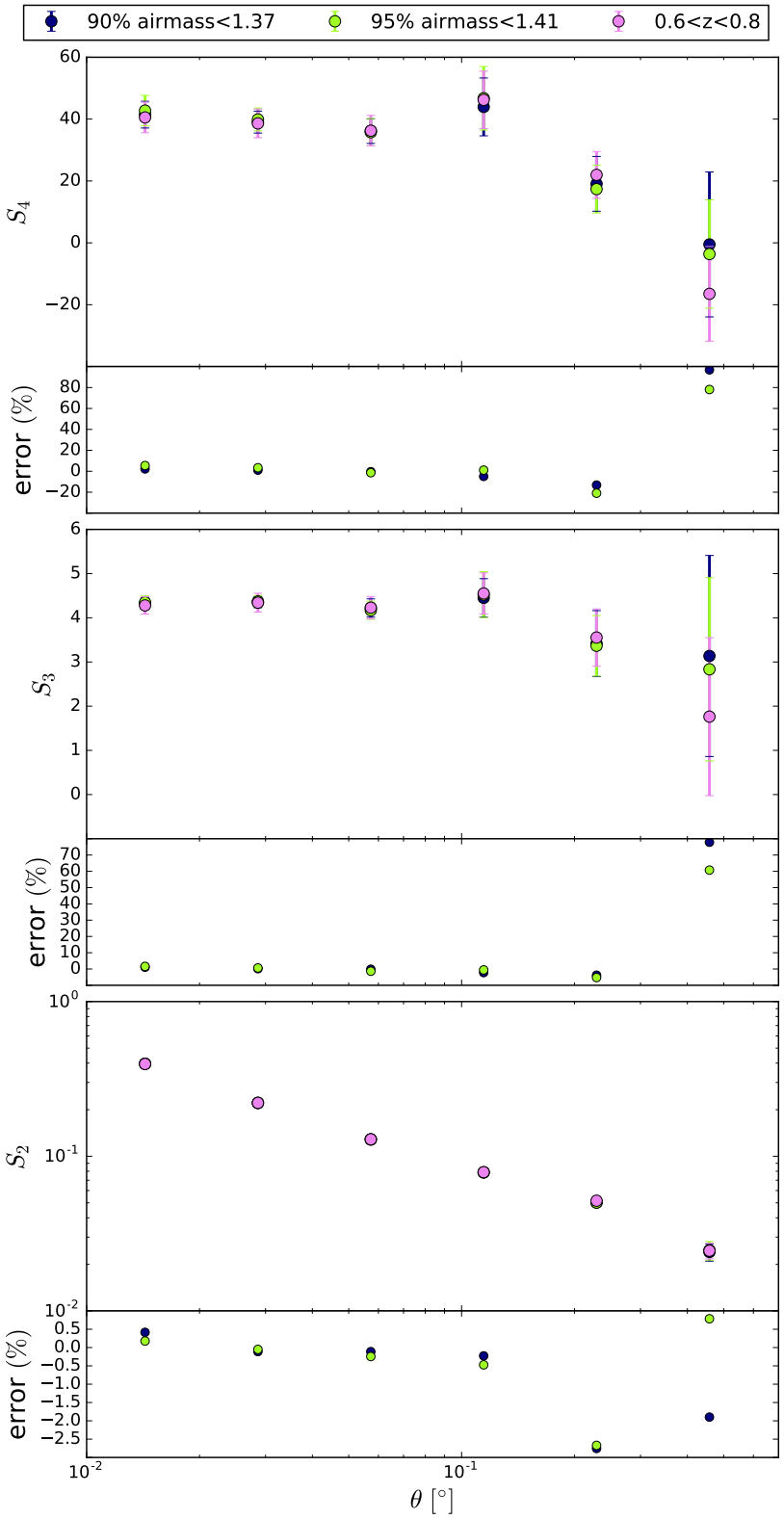


Figure A.12: Moments of the density contrast distribution of the DES SV benchmark sample for redshift bin $0.6 < z < 0.8$ and different airmass cuts.

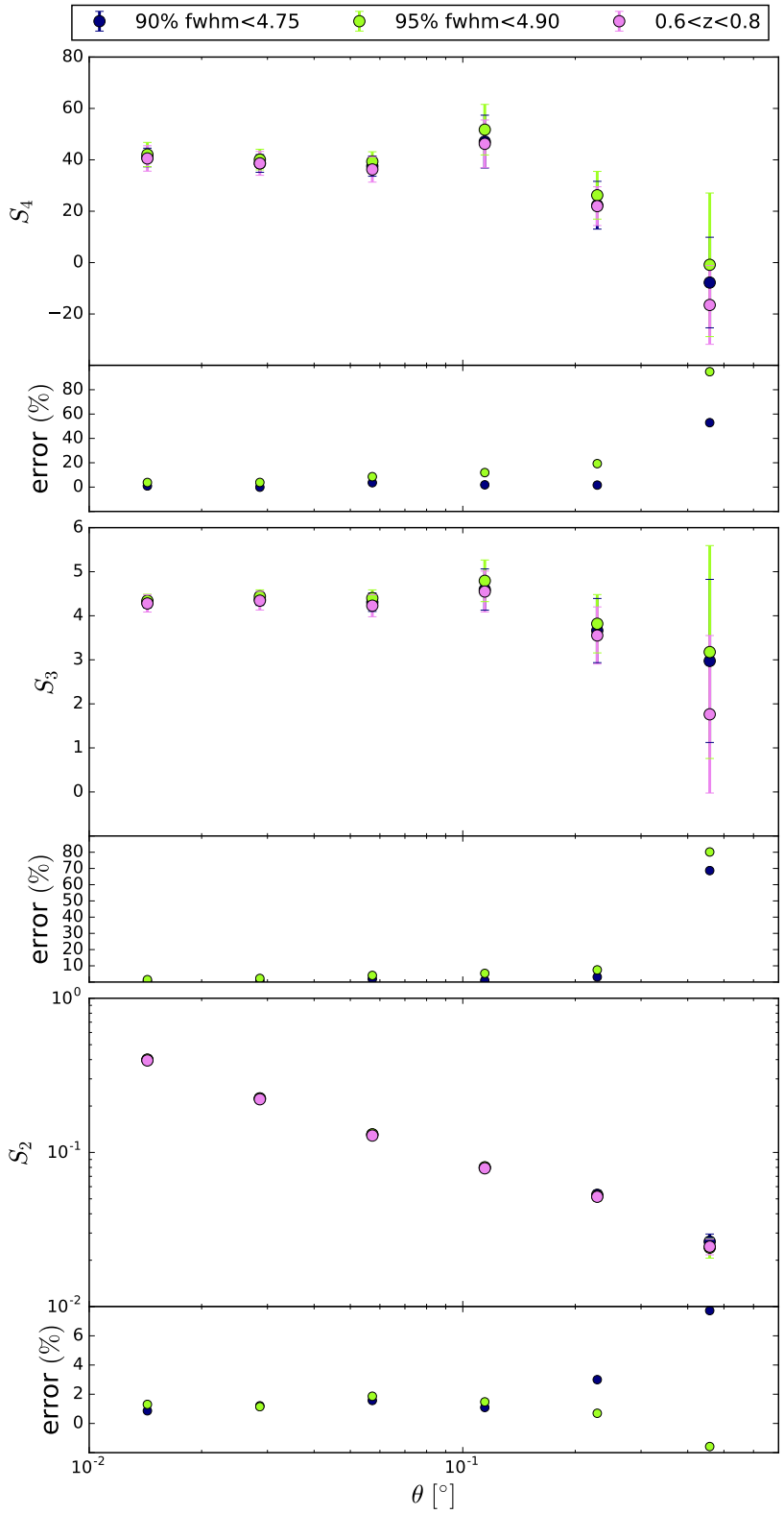


Figure A.13: Moments of the density contrast distribution of the DES SV benchmark sample for redshift bin $0.6 < z < 0.8$ and different seeing cuts.

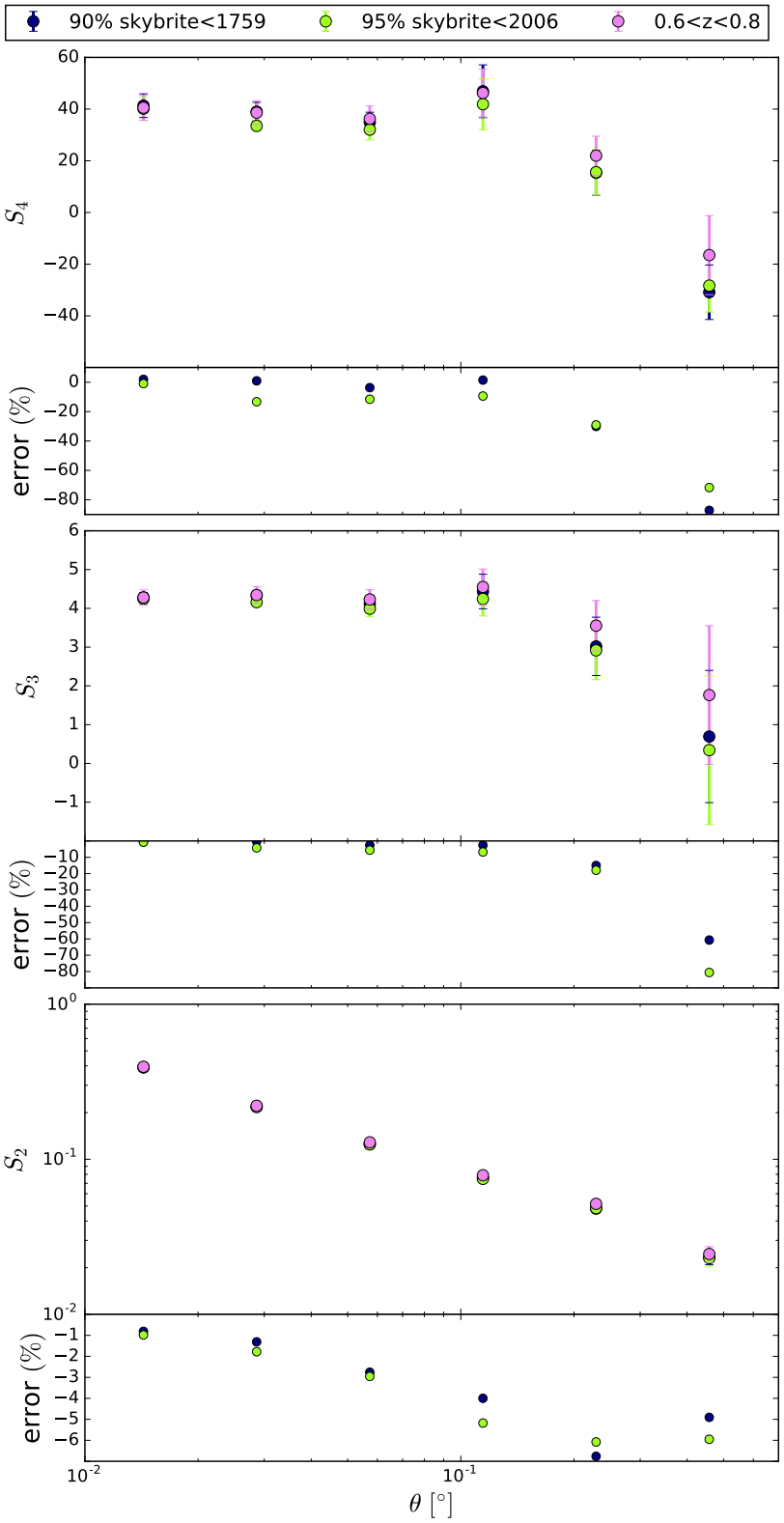


Figure A.14: Moments of the density contrast distribution of the DES SV benchmark sample for redshift bin $0.6 < z < 0.8$ and different skybrightness cuts.

Bibliography

[1] Michael J. Mortonson, David H. Weinberg, and Martin White. “Dark Energy: A Short Review”. In: (2013). arXiv: 1401.0046 [astro-ph.CO].

[2] Andreas Albrecht et al. “Report of the Dark Energy Task Force”. In: (2006). arXiv: astro-ph/0609591 [astro-ph].

[3] P. A. R. Ade et al. “Planck 2015 results. XIII. Cosmological parameters”. In: *Astron. Astrophys.* 594 (2016), A13. arXiv: 1502.01589 [astro-ph.CO].

[4] Albert Einstein. In: *Sitz. Preuss. Akad. Wiss. Phys.* 142 (1917).

[5] Steven Weinberg. *Gravitation and Cosmology*. John Wiley & Sons, 1972.

[6] Y. B. Zel’dovich, Andrzej Krasinski, and Y.B. Zeldovich. “The Cosmological constant and the theory of elementary particles”. In: *Sov. Phys. Usp.* 11 (1968). [Usp. Fiz. Nauk95,209(1968)], pp. 381–393.

[7] Edwin Hubble. “A relation between distance and radial velocity among extra-galactic nebulae”. In: *Proc. Nat. Acad. Sci.* 15 (1929), pp. 168–173.

[8] Adam G. Riess et al. “Observational evidence from supernovae for an accelerating universe and a cosmological constant”. In: *Astron. J.* 116 (1998), pp. 1009–1038. arXiv: astro-ph/9805201 [astro-ph].

[9] S. Perlmutter et al. “Measurements of Omega and Lambda from 42 high redshift supernovae”. In: *Astrophys. J.* 517 (1999), pp. 565–586. arXiv: astro-ph/9812133 [astro-ph].

[10] E.V. Linder. “Exploring the Expansion History of the Universe”. In: *Physical Review Letters* 90.9, 091301 (Mar. 2003), p. 091301.

[11] G. Gamow. “Expanding universe and the origin of elements”. In: *Phys. Rev.* 70 (1946), pp. 572–573.

[12] R. A. Alpher, H. Bethe, and G. Gamow. “The origin of chemical elements”. In: *Phys. Rev.* 73 (1948), pp. 803–804.

[13] R. A. Alpher and R. C. herman. “Remarks on the Evolution of the Expanding Universe”. In: *Phys. Rev.* 75 (1949), pp. 1089–1090.

[14] Juan Garcia-Bellido. “Cosmology and astrophysics”. In: *2004 European School of High-Energy Physics, Sant Feliu de Guixols, Spain, 30 May - 12 June 2004*. 2005, pp. 267–342. arXiv: astro-ph/0502139 [astro-ph].

[15] Alexei A. Starobinsky. “A New Type of Isotropic Cosmological Models Without Singularity”. In: *Phys. Lett.* B91 (1980), pp. 99–102.

[16] Alan H. Guth. “The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems”. In: *Phys. Rev.* D23 (1981), pp. 347–356.

[17] Andrei D. Linde. “A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems”. In: *Phys. Lett.* B108 (1982), pp. 389–393.

[18] Andreas Albrecht and Paul J. Steinhardt. “Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking”. In: *Phys. Rev. Lett.* 48 (1982), pp. 1220–1223.

[19] Andrew R. Liddle and D. H. Lyth. *Cosmological inflation and large scale structure*. 2000.

[20] Antonio Riotto. “Inflation and the theory of cosmological perturbations”. In: *Astroparticle physics and cosmology. Proceedings: Summer School, Trieste, Italy, Jun 17-Jul 5 2002*. 2002, pp. 317–413. arXiv: hep-ph/0210162 [hep-ph].

[21] Jerome Martin, Christophe Ringeval, and Vincent Vennin. “Encyclopædia Inflationaris”. In: *Phys. Dark Univ.* 5-6 (2014), pp. 75–235. arXiv: 1303.3787 [astro-ph.CO].

[22] D. J. Fixsen. “The Temperature of the Cosmic Microwave Background”. In: *Astrophys. J.* 707 (Dec. 2009), pp. 916–920. arXiv: 0911.1955.

[23] Martin Eriksen and Enrique Gaztanaga. “Combining spectroscopic and photometric surveys using angular cross-correlations III: Galaxy bias and stochastisity”. In: (2015). arXiv: 1508.00035 [astro-ph.CO].

[24] Richard H. Cyburt, Brian D. Fields, Keith A. Olive, and Tsung-Han Yeh. “Big Bang Nucleosynthesis: 2015”. In: *Rev. Mod. Phys.* 88 (2016), p. 015004. arXiv: 1505.01076 [astro-ph.CO].

[25] Sung-Chul Yoon and N. Langer. “Presupernova evolution of accreting white dwarfs with rotation”. In: *Astron. Astrophys.* 419 (2004), p. 623. arXiv: astro-ph/0402287 [astro-ph].

[26] P. A. R. Ade et al. “Planck 2015 results. XIII. Cosmological parameters”. In: *Astron. Astrophys.* 594 (2016), A13. arXiv: 1502.01589 [astro-ph.CO].

[27] Arno A. Penzias and Robert Woodrow Wilson. “A Measurement of excess antenna temperature at 4080-Mc/s”. In: *Astrophys. J.* 142 (1965), pp. 419–421.

[28] Stephane Courteau et al. “Shellflow. I. the convergence of the velocity field at 6000 km/s”. In: *Astrophys. J.* 544 (2000), pp. 636–640. arXiv: astro-ph/0002420 [astro-ph].

[29] George F. Smoot et al. “Structure in the COBE differential microwave radiometer first year maps”. In: *Astrophys. J.* 396 (1992), pp. L1–L5.

[30] N. Jarosik et al. “Three-year Wilkinson Microwave Anisotropy Pr (WMAP) observations: beam profiles, data processing, radiometer characterization and systematic error limits”. In: *Astrophys. J. Suppl.* 170 (2007), p. 263. arXiv: astro-ph/0603452 [astro-ph].

[31] G. Hinshaw et al. “Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Data Processing, Sky Maps, and Basic Results”. In: *Astrophys. J. Suppl.* 180 (2009), pp. 225–245. arXiv: 0803.0732 [astro-ph].

[32] P. A. R. Ade et al. “Detection of B-Mode Polarization at Degree Angular Scales by BICEP2”. In: *Phys. Rev. Lett.* 112.24 (2014), p. 241101. arXiv: 1403.3985 [astro-ph.CO].

[33] P. A. R. Ade et al. “Joint Analysis of BICEP2/Keck Array and Planck Data”. In: *Phys. Rev. Lett.* 114 (2015), p. 101301. arXiv: 1502.00612 [astro-ph.CO].

[34] P.J.E. Peebles and J.T. Yu. “Primeval adiabatic perturbation in an expanding universe”. In: *Astrophys.J.* 162 (1970), pp. 815–836.

[35] R.A. Sunyaev and Ya.B. Zeldovich. “Small scale fluctuations of relic radiation”. In: *Astrophys.Space Sci.* 7 (1970), pp. 3–19.

[36] P. de Bernardis et al. “A Flat universe from high resolution maps of the cosmic microwave background radiation”. In: *Nature* 404 (2000), pp. 955–959. arXiv: astro-ph/0004404 [astro-ph].

[37] Daniel J. Eisenstein et al. “Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies”. In: *Astrophys.J.* 633 (2005), pp. 560–574. arXiv: astro-ph/0501171 [astro-ph].

[38] Ashley J. Ross et al. “The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Observational systematics and baryon acoustic oscillations in the correlation function”. In: *Mon. Not. Roy. Astron. Soc.* 464.1 (2017), pp. 1168–1191. arXiv: 1607.03145 [astro-ph.CO].

[39] Florian Beutler et al. “The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in Fourier-space”. In: *Mon. Not. Roy. Astron. Soc.* 464.3 (2017), pp. 3409–3430–3430. arXiv: 1607.03149 [astro-ph.CO].

[40] Patrick McDonald and Daniel Eisenstein. “Dark energy and curvature from a future baryonic acoustic oscillation survey using the Lyman-α forest”. In: *Phys.Rev.* D76 (2007), p. 063009. arXiv: astro-ph/0607122 [astro-ph].

[41] Matthias Bartelmann and Peter Schneider. “Weak gravitational lensing”. In: *Phys. Rept.* 340 (2001), pp. 291–472. arXiv: astro-ph/9912508 [astro-ph].

[42] Alexandre Refregier. “Weak gravitational lensing by large scale structure”. In: *Ann. Rev. Astron. Astrophys.* 41 (2003), pp. 645–668. arXiv: astro-ph/0307212 [astro-ph].

[43] M. Kilbinger et al. “CFHTLenS: combined probe cosmological model comparison using 2D weak gravitational lensing”. In: *Mon. Not. Roy. Astron. Soc.* 430 (Apr. 2013), pp. 2200–2220. arXiv: 1212.3338.

[44] Catherine Heymans et al. “CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments”. In: *Mon. Not. Roy. Astron. Soc.* 432 (2013), p. 2433. arXiv: 1303.1808 [astro-ph.CO].

[45] H. Hildebrandt et al. “KiDS-450: Cosmological parameter constraints from tomographic weak gravitational lensing”. In: (2016). arXiv: 1606.05338 [astro-ph.CO].

[46] S. W. Allen, A. E. Evrard, and A. B. Mantz. “Cosmological Parameters from Observations of Galaxy Clusters”. In: *Annual Review of Astron and Astrophys* 49 (Sept. 2011), pp. 409–470. arXiv: 1103.4829 [astro-ph.CO].

[47] N. Kaiser. “Clustering in real space and in redshift space”. In: *Mon.Not.Roy.Astron.Soc.* 227 (1987), pp. 1–27.

[48] J.C. Jackson. “Fingers of God”. In: *Mon.Not.Roy.Astron.Soc.* 156 (1972), p. 1. arXiv: 0810.3908 [astro-ph].

[49] P. J. E. Peebles. *The large-scale structure of the universe*. 1980.

[50] Lado Samushia et al. “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measuring growth rate and geometry with anisotropic clustering”. In: *Mon. Not. Roy. Astron. Soc.* 439.4 (2014), pp. 3504–3519. arXiv: 1312.4899 [astro-ph.CO].

[51] Hong Guo et al. “Redshift-space clustering of SDSS galaxies – luminosity dependence, halo occupation distribution, and velocity bias”. In: *Mon. Not. Roy. Astron. Soc.* 453.4 (2015), pp. 4368–4383. arXiv: 1505.07861 [astro-ph.CO].

[52] Manuel Drees and Gilles Gerbier. “Dark Matter”. In: *Chin. Phys.* (2015).

[53] Juan García-Bellido. “Massive Primordial Black Holes as Dark Matter and their detection with Gravitational Waves”. In: *11th International LISA Symposium Zurich, Switzerland, September 5-9, 2016*. 2017. arXiv: 1702.08275 [astro-ph.CO].

[54] Jianglai Liu, Xun Chen, and Xiangdong Ji. “Current status of direct dark matter detection experiments”. In: *Nature Phys.* 13.3 (2017), pp. 212–216.

[55] Matthew Colless et al. “The 2dF Galaxy Redshift Survey: Spectra and redshifts”. In: *Mon.Not.Roy.Astron.Soc.* 328 (2001), p. 1039. arXiv: astro-ph/0106498 [astro-ph].

[56] D. Heath Jones et al. “The 6dF Galaxy Survey: Samples, observational techniques and the first data release”. In: *Mon.Not.Roy.Astron.Soc.* 355 (2004), pp. 747–763. arXiv: astro-ph/0403501 [astro-ph].

[57] Michael J. Drinkwater et al. “The WiggleZ Dark Energy Survey: Survey Design and First Data Release”. In: *Mon.Not.Roy.Astron.Soc.* 401 (2010), pp. 1429–1452. arXiv: 0911.4246 [astro-ph.CO].

[58] Christopher P. Ahn et al. “The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey”. In: *Astrophys.J.Suppl.* 203 (2012), p. 21. arXiv: 1207.7137 [astro-ph.IM].

[59] T. Abbott et al. “The dark energy survey”. In: (2005). arXiv: astro-ph/0510346 [astro-ph].

[60] Nicholas Kaiser et al. “Pan-STARRS: A Large Synoptic Survey Telescope Array”. In: *Proc.SPIE Int.Soc.Opt.Eng.* 4836 (2002), pp. 154–164.

[61] J. Anthony Tyson. “Large synoptic survey telescope: Overview”. In: *Proc.SPIE Int.Soc.Opt.Eng.* 4836 (2002), pp. 10–20. arXiv: astro-ph/0302102 [astro-ph].

[62] N. Benitez et al. “Measuring Baryon Acoustic Oscillations along the line of sight with photometric redshifs: the PAU survey”. In: *Astrophys.J.* 691 (2009), pp. 241–260. arXiv: 0807.0535 [astro-ph].

[63] N. Benitez et al. “J-PAS: The Javalambre-Physics of the Accelerated Universe Astrophysical Survey”. In: (2014). arXiv: 1403.5237 [astro-ph.CO].

[64] Pol Martí et al. “Precise photometric redshifts with a narrow-band filter set: The PAU Survey at the William Herschel Telescope”. In: *Mon. Not. Roy. Astron. Soc.* 442.1 (2014), pp. 92–109. arXiv: 1402.3220 [astro-ph.CO].

[65] Ofer Lahav and Andrew R Liddle. “The Cosmological Parameters 2014”. In: (2014). arXiv: 1401.1389 [astro-ph.CO].

[66] J. Kwan et al. “Cosmology from large scale galaxy clustering and galaxy-galaxy lensing with Dark Energy Survey Science Verification data”. In: *Mon. Not. Roy. Astron. Soc.* 464.4 (2017), pp. 4045–4062. arXiv: 1604.07871 [astro-ph.CO].

[67] Chris Blake et al. “The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations”. In: *Mon. Not. Roy. Astron. Soc.* 418 (2011), pp. 1707–1724. arXiv: 1108.2635 [astro-ph.CO].

[68] Adam G. Riess et al. “A 3% Solution: Determination of the Hubble Constant with the Hubble Space Telescope and Wide Field Camera 3”. In: *Astrophys.J.* 730 (2011), p. 119. arXiv: 1103.2976 [astro-ph.CO].

[69] LSST Science Collaboration, P.A. Abell, J. Allison, and S.F. Anderson. “LSST Science Book, Version 2.0”. In: *ArXiv e-prints* (Dec. 2009). arXiv: 0912.0201 [astro-ph.IM].

[70] Z. Ivezic et al. “LSST: from Science Drivers to Reference Design and Anticipated Data Products”. In: (2008). arXiv: 0805.2366 [astro-ph].

[71] R. Laureijs et al. “Euclid Definition Study Report”. In: *ArXiv e-prints* (Oct. 2011). arXiv: 1110.3193 [astro-ph.CO].

[72] Luca Amendola et al. “Cosmology and fundamental physics with the Euclid satellite”. In: *Living Rev. Rel.* 16 (2013), p. 6. arXiv: 1206.1225 [astro-ph.CO].

[73] Michael Levi et al. “The DESI Experiment, a whitepaper for Snowmass 2013”. In: (2013). arXiv: 1308.0847 [astro-ph.CO].

[74] D. J. Schlegel et al. “BigBOSS: The Ground-Based Stage IV Dark Energy Experiment”. In: *ArXiv e-prints* (Apr. 2009). arXiv: 0904.0468 [astro-ph.CO].

[75] F. Abdalla et al. “The Dark Energy Spectrometer (DESpec): A Multi-Fiber Spectroscopic Upgrade of the Dark Energy Camera and Survey for the Blanco Telescope”. In: *ArXiv e-prints* (Sept. 2012). arXiv: 1209.2451 [astro-ph.CO].

[76] Antony Lewis, Anthony Challinor, and Anthony Lasenby. “Efficient computation of CMB anisotropies in closed FRW models”. In: *Astrophys. J.* 538 (2000), pp. 473–476. arXiv: astro-ph/9911177 [astro-ph].

[77] Uros Seljak and Matias Zaldarriaga. “A Line of sight integration approach to cosmic microwave background anisotropies”. In: *Astrophys. J.* 469 (1996), pp. 437–444. arXiv: astro-ph/9603033 [astro-ph].

[78] J. Lesgourgues. “The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview”. In: *ArXiv e-prints* (Apr. 2011). arXiv: 1104.2932 [astro-ph.IM].

[79] Stephen D. Landy and Alexander S. Szalay. “Bias and variance of angular correlation functions”. In: *Astrophys.J.* 412 (1993), p. 64.

[80] D. Alonso. “CUTE solutions for two-point correlation functions from large cosmological datasets”. In: *ArXiv e-prints* (Oct. 2012). arXiv: 1210.1833 [astro-ph.IM].

[81] S. Dodelson. *Modern Cosmology*. Academic Press. ISBN: 9780122191411. Academic Press, 2003.

[82] James N. Fry and Robert J. Scherrer. “Skewness and nonGaussian initial conditions”. In: *Astrophys. J.* 429 (1994), pp. 36–42.

[83] James N. Fry and Enrique Gaztanaga. “Redshift distortions of galaxy correlation functions”. In: *Astrophys. J.* 425 (1994), pp. 1–13. arXiv: astro-ph/9305032 [astro-ph].

[84] M. Crocce and R. Scoccimarro. “Renormalized cosmological perturbation theory”. In: *PRD* 73.6, 063519 (Mar. 2006), p. 063519. eprint: astro-ph/0509418.

[85] R. Takahashi et al. “Revising the Halofit Model for the Nonlinear Matter Power Spectrum”. In: *Astrophysical Journal* 761, 152 (Dec. 2012), p. 152. arXiv: 1208.2701 [astro-ph.CO].

[86] Enrique Gaztañaga. “High order galaxy correlation functions in the APM galaxy survey”. In: *Mon. Not. Roy. Astron. Soc.* 268 (1994), p. 913. arXiv: astro-ph/9309019 [astro-ph].

[87] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro. “Large scale structure of the universe and cosmological perturbation theory”. In: *Phys. Rept.* 367 (2002), pp. 1–248. arXiv: astro-ph/0112551 [astro-ph].

[88] Francis Bernardeau. “Skewness and Kurtosis in large scale cosmic fields”. In: *Astrophys. J.* 433 (1994), p. 1. arXiv: astro-ph/9312026 [astro-ph].

[89] F. R. Bouchet, R. Juszkiewicz, S. Colombi, and R. Pellat. “Weakly nonlinear gravitational instability for arbitrary Omega”. In: *Astrophys. J.* 394 (1992), pp. L5–L8.

[90] Enrique Gaztanaga and J. Alberto Lobo. “Nonlinear gravitational growth of large scale structures inside and outside standard cosmology”. In: *Astrophys. J.* 548 (2001), pp. 47–59. arXiv: astro-ph/0003129 [astro-ph].

[91] M. Wolk et al. “Evolution of hierarchical clustering in the CFHTLS-Wide since z $\sim$ 1”. In: *Mon. Not. Roy. Astron. Soc.* 435 (Oct. 2013), pp. 2–17. arXiv: 1301.3301.

[92] Ashley J. Ross, Robert J. Brunner, and Adam D. Myers. “Precision Measurements of Higher-Order Angular Galaxy Correlations Using 11 Million SDSS Galaxies”. In: *Astrophys. J.* 649 (2006), pp. 48–62. arXiv: astro-ph/0605748 [astro-ph].

[93] Carlton M. Baugh et al. “The 2dF Galaxy Redshift Survey: Hierarchical galaxy clustering”. In: *Mon. Not. Roy. Astron. Soc.* 351 (2004), p. L44. arXiv: astro-ph/0401405 [astro-ph].

[94] William J. Frith, P. J. Outram, and T. Shanks. “High-order 2MASS galaxy correlation functions: Probing the primordial density field and the linearity of galaxy bias”. In: *Mon. Not. Roy. Astron. Soc.* 373 (2006), pp. 759–768. arXiv: astro-ph/0507704 [astro-ph].

[95] C. Marinoni et al. “The VIMOS VLT Deep Survey: Testing the gravitational instability paradigm at z $\sim$1”. In: *Astron. Astrophys.* 487 (2008), p. 7. arXiv: 0802.1838 [astro-ph].

[96] Enrique Gaztañaga. “Galaxy clustering in the Sloan Digital Sky Survey (SDSS): A First comparison with the APM galaxy survey”. In: *Mon. Not. Roy. Astron. Soc.* 333 (2002), p. L21. arXiv: astro-ph/0106379 [astro-ph].

[97] Abel Yang and William C. Saslaw. “The galaxy counts-in-cells distribution from the SDSS”. In: *Astrophys. J.* 729 (2011), p. 123. arXiv: 1009.0013 [astro-ph.CO].

[98] W.C. Saslaw and P. Crane. “The scale dependence of galaxy distribution functions”. In: *Astrophys. J.* 380 (Oct. 1991), pp. 315–319.

[99] Peter Coles and Bernard Jones. “A Lognormal model for the cosmological mass distribution”. In: *Mon. Not. Roy. Astron. Soc.* 248 (1991), pp. 1–13.

[100] Simon D. M. White and M. J. Rees. “Core condensation in heavy halos: A Two stage theory for galaxy formation and clusters”. In: *Mon. Not. Roy. Astron. Soc.* 183 (1978), pp. 341–358.

[101] M. Davis, M.J. Geller, and J. Huchra. “The local mean mass density of the universe - New methods for studying galaxy clustering”. In: *Astrophys. J.* 221 (Apr. 1978), pp. 1–18.

[102] A. Dressler. “Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies”. In: *Astrophys. J.* 236 (Mar. 1980), pp. 351–365.

[103] Nick Kaiser. “On the Spatial correlations of Abell clusters”. In: *Astrophys. J.* 284 (1984), pp. L9–L12.

[104] J.M. Bardeen, J.R. Bond, N. Kaiser, and A.S. Szalay. “The statistics of peaks of Gaussian random fields”. In: *Astrophys. J.* 304 (May 1986), pp. 15–61.

[105] James N. Fry and Enrique Gaztanaga. “Biasing and hierarchical statistics in large scale structure”. In: *Astrophys. J.* 413 (1993), pp. 447–452. arXiv: astro-ph/9302009 [astro-ph].

[106] F. Bernardeau. “The Large scale gravitational bias from the quasilinear regime”. In: *Astron. Astrophys.* 312 (1996), pp. 11–23. arXiv: astro-ph/9602072 [astro-ph].

[107] H. J. Mo and Simon D. M. White. “An Analytic model for the spatial clustering of dark matter halos”. In: *Mon. Not. Roy. Astron. Soc.* 282 (1996), p. 347. arXiv: astro-ph/9512127 [astro-ph].

[108] Ravi K. Sheth and Giuseppe Tormen. “Large scale bias and the peak background split”. In: *Mon. Not. Roy. Astron. Soc.* 308 (1999), p. 119. arXiv: astro-ph/9901122 [astro-ph].

[109] Marc Manera, Ravi K Sheth, and Roman Scoccimarro. “Large scale bias and the inaccuracy of the peak-background split”. In: *Mon. Not. Roy. Astron. Soc.* 402 (2010), p. 589. arXiv: 0906.1314 [astro-ph.CO].

[110] M. Manera and E. Gaztañaga. “The local bias model in the large-scale halo distribution”. In: *Mon. Not. Roy. Astron. Soc.* 415 (July 2011), pp. 383–398. arXiv: 0912.0446.

[111] W.H. Press and P. Schechter. “Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation”. In: *Astrophys. J.* 187 (Feb. 1974), pp. 425–438.

[112] Adi Nusser and Marc Davis. “On the prediction of velocity fields from redshift space galaxy samples”. In: *Astrophys. J.* 421 (1994), pp. L1–L4. arXiv: astro-ph/9309009 [astro-ph].

[113] Max Tegmark and P. J. E. Peebles. “The Time evolution of bias”. In: *Astrophys. J.* 500 (1998), p. L79. arXiv: astro-ph/9804067 [astro-ph].

[114] M. Blanton et al. “Time Evolution of Galaxy Formation and Bias in Cosmological Simulations”. In: *Astrophys. J.* 531 (Mar. 2000), pp. 1–16. eprint: astro-ph/9903165.

[115] Arnau Pujol, Kai Hoffmann, Noelia Jiménez, and Enrique Gaztañaga. “What determines large scale galaxy clustering: halo mass or local density?” In: *Astron. Astrophys.* 598 (2017), A103. arXiv: 1510.01692 [astro-ph.CO].

[116] C. N. A. Willmer et al. “Two galaxy clusters: a3565 and a3560”. In: *Astron. J.* 118 (1999), p. 1131. arXiv: astro-ph/9906023 [astro-ph].

[117] I. Zehavi et al. “Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data”. In: *Astrophys. J.* 571 (May 2002), pp. 172–190. eprint: astro-ph/0106476.

[118] P. Norberg, C.M. Baugh, E. Hawkins, and S. Maddox. “The 2dF Galaxy Redshift Survey: the dependence of galaxy clustering on luminosity and spectral type”. In: *Mon. Not. Roy. Astron. Soc.* 332 (June 2002), pp. 827–838. eprint: astro-ph/0112043.

[119] E. Gaztańaga et al. “Cross-Correlation of spectroscopic and photometric galaxy surveys: cosmology from lensing and redshift distortions”. In: (2011). arXiv: 1109.4852 [astro-ph.CO].

[120] J. N. Fry. “The Evolution of Bias”. In: *The Astrophysical Journal Letters* 461.2 (1996), p. L65.

[121] Robert J. Scherrer and David H. Weinberg. “Constraints on the effects of locally-biased galaxy formation”. In: *Astrophys. J.* 504 (1998), pp. 607–611. arXiv: astro-ph/9712192 [astro-ph].

[122] J.A. Frieman and E. Gaztañaga. “The Projected Three-Point Correlation Function: Theory and Observations”. In: *Astrophys. J.l* 521 (Aug. 1999), pp. L83–L86. eprint: astro-ph/9903423.

[123] Ashley J. Ross, Robert J. Brunner, and Adam D. Myers. “Higher-Order Angular Galaxy Correlations in the SDSS: Redshift and Color Dependence of non-Linear Bias”. In: *Astrophys. J.* 665 (2007), pp. 67–84. arXiv: 0704.2573 [astro-ph].

[124] Julien Bel, Kai Hoffmann, and Enrique Gaztañaga. “Non-local bias contribution to third-order galaxy correlations”. In: *Mon. Not. Roy. Astron. Soc.* 453.1 (2015), pp. 259–276. arXiv: 1504.02074 [astro-ph.CO].

[125] Jennifer E. Pollack, Robert E. Smith, and Cristiano Porciani. “A new method to measure galaxy bias”. In: *Mon. Not. Roy. Astron. Soc.* 440.1 (2014), pp. 555–576. arXiv: 1309.0504 [astro-ph.CO].

[126] Kai Hoffmann, Julien Bel, and Enrique Gaztanaga. “Comparing halo bias from abundance and clustering”. In: *Mon. Not. Roy. Astron. Soc.* 450.2 (2015), pp. 1674–1692. arXiv: 1503.00313 [astro-ph.CO].

[127] Kwan Chuen Chan, Roman Scoccimarro, and Ravi K. Sheth. “Gravity and Large-Scale Non-local Bias”. In: *Phys. Rev.* D85 (2012), p. 083509. arXiv: 1201.3614 [astro-ph.CO].

[128] T. Baldauf, U. Seljak, V. Desjacques, and P. McDonald. “Evidence for quadratic tidal tensor bias from the halo bispectrum”. In: *Physical Review D* 86.8, 083540 (Oct. 2012), p. 083540. arXiv: 1201.4827 [astro-ph.CO].

[129] S. Cole and N. Kaiser. “Biased clustering in the cold dark matter cosmogony”. In: *Mon. Not. Roy. Astron. Soc.* 237 (Apr. 1989), pp. 1127–1146.

[130] A. Faltenbacher and S. D. M. White. “Assembly Bias and the Dynamical Structure of Dark Matter Halos”. In: *Astrophys. J.* 708 (Jan. 2010), pp. 469–473. arXiv: 0909.4302.

[131] Arnau Pujol and Enrique Gaztañaga. “Are the halo occupation predictions consistent with large scale galaxy clustering?” In: *Mon. Not. Roy. Astron. Soc.* 442.3 (2014), pp. 1930–1941. arXiv: 1306.5761 [astro-ph.CO].

[132] Arnau Pujol et al. “A new method to measure galaxy bias by combining the density and weak lensing fields”. In: *Mon. Not. Roy. Astron. Soc.* 462.1 (2016), pp. 35–47. arXiv: 1601.00160 [astro-ph.CO].

[133] Anatoly A. Klypin, Andrey V. Kravtsov, Octavio Valenzuela, and Francisco Prada. “Where are the missing Galactic satellites?” In: *Astrophys. J.* 522 (1999), pp. 82–92. arXiv: astro-ph/9901240 [astro-ph].

[134] N. Roth and C. Porciani. “Testing standard perturbation theory and the Eulerian local biasing scheme against N-body simulations”. In: *Mon. Not. Roy. Astron. Soc.* 415 (July 2011), pp. 829–844. arXiv: 1101.1520.

[135] I. Zehavi, Z. Zheng, D.H. Weinberg, and M.R. Blanton. “Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity”. In: *Astrophys. J.* 736, 59 (July 2011), p. 59. arXiv: 1005.2413.

[136] M. Cacciato et al. “On combining galaxy clustering and weak lensing to unveil galaxy biasing via the halo model”. In: *Mon. Not. Roy. Astron. Soc.* 426 (Oct. 2012), pp. 566–587. arXiv: 1203.2616.

[137] J. Coupon et al. “Galaxy clustering in the CFHTLS-Wide: the changing relationship between galaxies and haloes since z$\sim$1.2”. In: *Astronomy and Astrophysics* 542, A5 (June 2012), A5. arXiv: 1107.0616.

[138] E. Jullo et al. “COSMOS: Stochastic Bias from Measurements of Weak Lensing and Galaxy Clustering”. In: *Astrophys. J.* 750, 37 (May 2012), p. 37. arXiv: 1202.6491.

[139] Felipe A. Marin et al. “The WiggleZ Dark Energy Survey: constraining galaxy bias and cosmic growth with 3-point correlation functions”. In: *Mon. Not. Roy. Astron. Soc.* 432 (2013), p. 2654. arXiv: 1303.6644 [astro-ph.CO].

[140] C. Di Porto et al. “The VIMOS Public Extragalactic Redshift Survey (VIPERS) - Measuring non-linear galaxy bias at z∼0.8”. In: *Astron. Astrophys.* 594 (2016), A62. arXiv: 1406.6692 [astro-ph.CO].

[141] A. Durkalec, O. Le Fèvre, A. Pollo, and S. de la Torre. “Evolution of clustering length, large-scale bias, and host halo mass at 2<z<5 in the VIMOS Ultra Deep Survey (VUDS)”. In: *Astronomy and Astrophysics* 583, A128 (Nov. 2015), A128. arXiv: 1411.5688.

[142] M. Crocce et al. “Galaxy clustering, photometric redshifts and diagnosis of systematics in the DES Science Verification data”. In: *MNRAS* 455 (Feb. 2016), pp. 4301–4324. arXiv: 1507.05360.

[143] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. “A Universal density profile from hierarchical clustering”. In: *Astrophys. J.* 490 (1997), pp. 493–508. arXiv: astro-ph/9611107 [astro-ph].

[144] Zheng Zheng et al. “Theoretical models of the halo occupation distribution: Separating central and satellite galaxies”. In: *Astrophys. J.* 633 (2005), pp. 791–809. arXiv: astro-ph/0408564 [astro-ph].

[145] Michael J. I. Brown et al. “Red Galaxy Growth and the Halo Occupation Distribution”. In: *Astrophys. J.* 682 (2008), pp. 937–963. arXiv: 0804.2293 [astro-ph].

[146] J.L. Tinker et al. “The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests”. In: *Astrophys. J.* 724 (Dec. 2010), pp. 878–886. arXiv: 1001.3162.

[147] T. Abbott et al. “The Dark Energy Survey: more than dark energy – an overview”. In: *Mon. Not. Roy. Astron. Soc.* 460.2 (2016), pp. 1270–1299. arXiv: 1601.00329 [astro-ph.CO].

[148] H. T. Diehl. “The Dark Energy Survey and Operations: Years 1 to 3”. In: 2016.

[149] B. Flaugher, Diehl, and DES Collaboration. “The Dark Energy Camera”. In: *Astronomical Journal* 150, 150 (Nov. 2015), p. 150. arXiv: 1504.02900 [astro-ph.IM].

[150] D.W. Gerdes et al. “Observation of Two New L4 Neptune Trojans in the Dark Energy Survey Supernova Fields”. In: *Astronomical Journal* 151, 39 (Feb. 2016), p. 39. arXiv: 1507.05177 [astro-ph.EP].

[151] T.S. Li et al. “Discovery of a Stellar Overdensity in Eridanus-Phoenix in the Dark Energy Survey”. In: *Astrophys. J.* 817, 135 (Feb. 2016), p. 135. arXiv: 1509.04296.

[152] K. Bechtol et al. “Eight New Milky Way Companions Discovered in First-year Dark Energy Survey Data”. In: *Astrophys. J.* 807, 50 (July 2015), p. 50. arXiv: 1503.02584.

[153] A. Drlica-Wagner et al. “Eight Ultra-faint Galaxy Candidates Discovered in Year Two of the Dark Energy Survey”. In: *Astrophys. J.* 813.2 (2015), p. 109. arXiv: 1508.03622 [astro-ph.GA].

[154] Y. Zhang et al. “Galaxies in X-ray Selected Clusters and Groups in Dark Energy Survey Data I: Stellar Mass Growth of Bright Central Galaxies Since z$\sim$1.2”. In: *Astrophys. J.* 816.2 (2016), p. 98. arXiv: 1504.02983 [astro-ph.GA].

[155] C. Chang et al. “Wide-Field Lensing Mass Maps from Dark Energy Survey Science Verification Data”. In: *Phys. Rev. Lett.* 115.5 (2015), p. 051301. arXiv: 1505.01871 [astro-ph.CO].

[156] D. Gruen et al. “Weak lensing by galaxy troughs in DES Science Verification data”. In: *Mon. Not. Roy. Astron. Soc.* 455.3 (2016), pp. 3367–3380. arXiv: 1507.05090 [astro-ph.CO].

[157] T. Abbott et al. “Cosmology from cosmic shear with Dark Energy Survey Science Verification data”. In: *Phys. Rev.* D94.2 (2016), p. 022001. arXiv: 1507.05552 [astro-ph.CO].

[158] M. Soares-Santos et al. “A Dark Energy Camera Search for an Optical Counterpart to the First Advanced LIGO Gravitational Wave Event GW150914”. In: *Astrophys. J.* 823.2 (2016), p. L33. arXiv: 1602.04198 [astro-ph.CO].

[159] J. Annis et al. “A Dark Energy Camera Search for Missing Supergiants in the LMC After the Advanced LIGO Gravitational Wave Event GW150914”. In: *Astrophys. J.* 823.2 (2016), p. L34. arXiv: 1602.04199 [astro-ph.HE].

[160] M. Lueker et al. “Measurements of Secondary Cosmic Microwave Background Anisotropies with the South Pole Telescope”. In: *Astrophys. J.* 719 (Aug. 2010), pp. 1045–1066. arXiv: 0912.4317.

[161] C.P. Ahn et al. “The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey”. In: *Astrophys. J.s* 203, 21 (Dec. 2012), p. 21. arXiv: 1207.7137 [astro-ph.IM].

[162] J.J. Mohr et al. “The Dark Energy Survey data processing and calibration system”. In: *Software and Cyberinfrastructure for Astronomy II*. Vol. 8451. Proceedings of the SPIE. Sept. 2012, p. 84510D. arXiv: 1207.3189 [astro-ph.IM].

[163] C. E. Cunha et al. “Spectroscopic failures in photometric redshift calibration: cosmological biases and survey requirements”. In: *Mon. Not. Roy. Astron. Soc.* 444 (Oct. 2014), pp. 129–146. arXiv: 1207.3347.

[164] C. Sanchez et al. “Photometric redshift analysis in the Dark Energy Survey Science Verification data”. In: *Mon. Not. Roy. Astron. Soc.* 445.2 (2014), pp. 1482–1506. arXiv: 1406.4407 [astro-ph.IM].

[165] Narciso Benitez. “Bayesian photometric redshift estimation”. In: *Astrophys.J.* 536 (2000), pp. 571–583. arXiv: astro-ph/9811189 [astro-ph].

[166] M. Carrasco Kind and R. J. Brunner. “TPZ: photometric redshift PDFs and ancillary information by using prediction trees and random forests”. In: *MNRAS* 432 (June 2013), pp. 1483–1501. arXiv: 1303.7269.

[167] E. Bertin and S. Arnouts. “SExtractor: Software for source extraction.” In: *Astronomy and Astrophysics, Supplement* 117 (June 1996), pp. 393–404.

[168] M. T. Busha et al. “Catalog Production for the DES Blind Cosmology Challenge”. In: *American Astronomical Society Meeting Abstracts #221*. Vol. 221. American Astronomical Society Meeting Abstracts. Jan. 2013, p. 341.07.

[169] C. Chang et al. “Modeling the Transfer Function for the Dark Energy Survey”. In: *Astrophys. J.* 801.2 (2015), p. 73. arXiv: 1411.0032 [astro-ph.IM].

[170] B. Leistedt et al. “Mapping and simulating systematics due to spatially-varying observing conditions in DES Science Verification data”. In: *Astrophys. J. Suppl.* 226.2 (2016), p. 24. arXiv: 1507.05647 [astro-ph.CO].

[171] M. R. Becker et al. “Cosmic shear measurements with Dark Energy Survey Science Verification data”. In: *Phys. Rev.* D94.2 (2016), p. 022002. arXiv: 1507.05598 [astro-ph.CO].

[172] J. Clampitt et al. “Galaxy–galaxy lensing in the Dark Energy Survey Science Verification data”. In: *Mon. Not. Roy. Astron. Soc.* 465.4 (2017), pp. 4204–4218. arXiv: 1603.05790 [astro-ph.CO].

[173] P. Fosalba, E. Gaztañaga, F. J. Castander, and M. Manera. “The onion universe: all sky lightcone simulations in spherical shells”. In: *MNRAS* 391 (Nov. 2008), pp. 435–446. arXiv: 0711.1540.

[174] M. Crocce, P. Fosalba, F. J. Castander, and E. Gaztañaga. “Simulating the Universe with MICE: the abundance of massive clusters”. In: *MNRAS* 403 (Apr. 2010), pp. 1353–1367. arXiv: 0907.0019.

[175] M. Crocce et al. “The MICE Grand Challenge Lightcone Simulation II: Halo and Galaxy catalogues”. In: *Mon. Not. Roy. Astron. Soc.* 453 (2015), pp. 1513–1530. arXiv: 1312.2013 [astro-ph.CO].

[176] P.A.R. Ade et al. “Planck 2013 results. XVI. Cosmological parameters”. In: (2013). arXiv: 1303.5076 [astro-ph.CO].

[177] K. M. Górski et al. “HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere”. In: *ApJ* 622 (Apr. 2005), pp. 759–771. eprint: astro-ph/0409513.

[178] T. Giannantonio et al. “CMB lensing tomography with the DES Science Verification galaxies”. In: *MNRAS* 456 (Mar. 2016), pp. 3213–3244. arXiv: 1507.05551.

[179] J. Prat et al. “Galaxy bias from galaxy-galaxy lensing in the DES Science Verification Data”. In: *ArXiv e-prints* (Sept. 2016). arXiv: 1609.08167.

[180] B. Leistedt et al. “Mapping and Simulating Systematics due to Spatially Varying Observing Conditions in DES Science Verification Data”. In: *The Astrophysical Journal Suplement* 226, 24 (Oct. 2016), p. 24. arXiv: 1507.05647.

[181] D.G. Monet et al. “The USNO-B Catalog”. In: *The Astrophysical Journal* 125 (Feb. 2003), pp. 984–993. eprint: astro-ph/0210694.

[182] D.J. Schlegel, D.P. Finkbeiner, and M. Davis. “Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds”. In: *The Astrophysical Journal* 500 (June 1998), pp. 525–553. eprint: astro-ph/9710327.

[183] L. Clerkin et al. “Testing the lognormality of the galaxy and weak lensing convergence distributions from Dark Energy Survey maps”. In: *Mon. Not. Roy. Astron. Soc.* (2016). arXiv: 1605.02036 [astro-ph.CO].